

GenCore version 5.1.4_p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 27, 2003, 15:10:53 ; Search time 41 Seconds
(without alignments)
1514,507 Million cell updates/sec

Title: US-09-934-066-2
Perfect score: 2454
Sequence: 1 MSSPLGHFQILVFLHALLIF.....CNMGVDVKQTVAIEQACSM 466

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002.*
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2454	100.0	466	21	AA2002
2	2446	99.7	466	21	AA2001
3	2446	99.7	485	21	AA2000
4	2009	81.9	380	21	AA2002
5	2001	81.5	380	21	AA2001
6	1554	63.3	295	21	AA2002
7	1246	50.8	440	14	AA2001
8	1246	50.8	490	21	AA2002
9	1246	50.8	494	21	AA2001
10	1245	50.7	440	14	AA2002

11	1215	49.5	440	14	AA2002
12	1080	44.0	440	14	AA2001
13	1023	41.7	395	21	AA2000
14	828	33.7	433	19	AA2002
15	828	33.7	433	19	AA2001
16	771	31.4	376	19	AA2002
17	749	30.5	363	14	AA2001
18	744	30.3	431	18	AA2002
19	744	30.3	431	20	AA2001
20	744	30.3	431	21	AA2002
21	685	27.9	435	23	AA2001
22	613	25.0	318	14	AA2002
23	505.5	20.6	414	22	AA2001
24	326.5	13.3	322	22	AA2002
25	269	11.0	381	21	AA2001
26	269	11.0	381	21	AA2002
27	254.5	10.4	383	21	AA2001
28	254.5	10.4	390	21	AA2002
29	216.5	8.8	182	22	AA2001
30	209.5	8.5	345	21	AA2002
31	209	8.5	345	21	AA2001
32	162.5	6.6	144	21	AA2002
33	141.5	5.8	1003	23	AA2001
34	141.5	5.8	1014	23	AA2002
35	137	5.6	619	23	AA2001
36	114.5	4.7	580	20	AA2002
37	112.5	4.6	1512	22	AA2001
38	112.5	4.6	1512	22	AA2002
39	112.5	4.6	1550	22	AA2001
40	109	4.4	507	21	AA2002
41	108.5	4.4	885	23	AA2001
42	108.5	4.4	885	23	AA2002
43	108	4.4	679	21	AA2001
44	107	4.4	996	22	AA2002
45	104.5	4.3	1651	21	AA2001

ALIGNMENTS

10261

RESULT 1
AAG52573
ID AAG52573 standard; Protein: 466 AA.

AC AAG52573;

DT 18-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 66847.

KW Protein identification; signal transduction pathway; metabolic pathway; hybridisation assay; genetic mapping; gene expression control; promoter; termination sequence.

OS Arabidopsis thaliana.

PN EPI033405-A2.

PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

PR 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 23-MAR-1999; 99US-0123548.

PR 25-MAR-1999; 99US-0125788.

PR 29-MAR-1999; 99US-0126264.

PR 01-APR-1999; 99US-0126785.

PR 06-APR-1999; 99US-0127462.

PR 08-APR-1999; 99US-0128234.

PR 16-APR-1999; 99US-0128714.

PR 19-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130591.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-01343768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138340.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141642.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146398.
PR 02-AUG-1999; 99US-0146399.
PR 03-AUG-1999; 99US-0147036.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151328.
PR 01-SEP-1999; 99US-0151320.
PR 07-SEP-1999; 99US-0152593.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 06-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.

[illegible]

```

PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0143390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 27-JUL-1999; 99US-0145951.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 10-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149368.
PR 18-AUG-1999; 99US-0149175.
PR 20-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 23-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151436.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.

```

```

PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 13-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 22-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160981.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

Query Match 99.7%; Score 2446; DB 21; Length 466;

Best Local Similarity 99.8%; Pred. No. 3e-209;

Matches 465; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

QY 1 MSSPLGHFQILVFLHALLIFSASERKTQLLNDNDVDESSDKSAGTRWAVLVAGSNEYNY 60
DB 1 MSSPLGHFQILVFLHALLIFSASERKTQLLNDNDVDESSDKSAGTRWAVLVAGSNEYNY 60
QY 61 RHQADICHAYQILRKGLKXENIIVFMYDDIAFSSSENPRPGVIIINPKDGEDVYKGVPKDY 120
DB 61 RHQADICHAYQILRKGLKXENIIVFMYDDIAFSSSENPRPGVIIINPKDGEDVYKGVPKDY 120
QY 121 TKEAVNVQNFYNVLLGNESGVTGGNGKVVKSGPNDNFIYYADHGAAPGLIAMPTGDEYMA 180
DB 121 TKEAVNVQNFYNVLLGNESGVTGGNGKVVKSGPNDNFIYYADHGAAPGLIAMPTGDEYMA 180
QY 181 KDFNEVLEKMHKRYKXKVIYVEACESGMFEGILKKNLIYAVTAANSKRESSNGVYCP 240
DB 181 KDFNEVLEKMHKRYKXKVIYVEACESGMFEGILKKNLIYAVTAANSKRESSNGVYCP 240
QY 241 ESYPPPESEIGTCGLDFTSISWLESDLDHMSKETLEOQYHVVKRVSDDVPTSHVCRF 300
DB 241 ESYPPPESEIGTCGLDFTSISWLESDLDHMSKETLEOQYHVVKRVSDDVPTSHVCRF 300
QY 301 GTEKMLKDYLSYIGRNPNNDNFTFTFESSPISNSGLVNPDRDIPLLYLQRIQAPMCS 360
DB 301 GTEKMLKDYLSYIGRNPNNDNFTFTFESSPISNSGLVNPDRDIPLLYLQRIQAPMCS 360

```

PR	14-JUN-1999;	99US-0139119;
PR	16-JUN-1999;	99US-0139452;
PR	16-JUN-1999;	99US-0139453;
PR	17-JUN-1999;	99US-0139492;
PR	17-JUN-1999;	99US-0139454;
PR	18-JUN-1999;	99US-0139455;
PR	18-JUN-1999;	99US-0139456;
PR	18-JUN-1999;	99US-0139457;
PR	18-JUN-1999;	99US-0139458;
PR	18-JUN-1999;	99US-0139459;
PR	18-JUN-1999;	99US-0139460;
PR	18-JUN-1999;	99US-0139461;
PR	18-JUN-1999;	99US-0139462;
PR	18-JUN-1999;	99US-0139463;
PR	18-JUN-1999;	99US-0139750;
PR	18-JUN-1999;	99US-0139763;
PR	21-JUN-1999;	99US-0139817;
PR	21-JUN-1999;	99US-0139899;
PR	22-JUN-1999;	99US-0140355;
PR	23-JUN-1999;	99US-0140354;
PR	23-JUN-1999;	99US-0140355;
PR	24-JUN-1999;	99US-0140695;
PR	28-JUN-1999;	99US-0140823;
PR	29-JUN-1999;	99US-0140991;
PR	30-JUN-1999;	99US-0141287;
PR	01-JUL-1999;	99US-0141842;
PR	02-JUL-1999;	99US-0142154;
PR	02-JUL-1999;	99US-0142055;
PR	15-JUL-1999;	99US-0144331;
PR	15-JUL-1999;	99US-0144362;
PR	15-JUL-1999;	99US-0144005;
PR	08-AUG-1999;	99US-0142803;
PR	09-JUL-1999;	99US-0142920;
PR	12-JUL-1999;	99US-0142977;
PR	13-JUL-1999;	99US-0143542;
PR	14-JUL-1999;	99US-0144332;
PR	19-JUL-1999;	99US-0144333;
PR	19-JUL-1999;	99US-0144332;
PR	19-JUL-1999;	99US-0144333;
PR	19-JUL-1999;	99US-0144334;
PR	19-JUL-1999;	99US-0144335;
PR	20-JUL-1999;	99US-0144352;
PR	20-JUL-1999;	99US-0144632;
PR	21-JUL-1999;	99US-0144884;
PR	21-JUL-1999;	99US-0144814;
PR	21-JUL-1999;	99US-0145086;
PR	21-JUL-1999;	99US-0145088;
PR	21-JUL-1999;	99US-0145085;
PR	22-JUL-1999;	99US-0145087;
PR	22-JUL-1999;	99US-0145089;
PR	22-JUL-1999;	99US-0145192;
PR	23-JUL-1999;	99US-0145145;
PR	23-JUL-1999;	99US-0145919;
PR	23-JUL-1999;	99US-0145218;
PR	23-JUL-1999;	99US-0145224;
PR	26-JUL-1999;	99US-0145276;
PR	27-JUL-1999;	99US-0145913;
PR	27-JUL-1999;	99US-0145919;
PR	27-JUL-1999;	99US-0147204;
PR	02-AUG-1999;	99US-0146386;
PR	02-AUG-1999;	99US-0146388;
PR	02-AUG-1999;	99US-0146389;
PR	03-AUG-1999;	99US-0147038;
PR	04-AUG-1999;	99US-0147202;
PR	04-AUG-1999;	99US-0147302;
PR	05-AUG-1999;	99US-0147192;
PR	05-AUG-1999;	99US-0147260;
PR	06-AUG-1999;	99US-0147303;
PR	06-AUG-1999;	99US-0147416;
PR	09-AUG-1999;	99US-0147493;
PR	10-AUG-1999;	99US-0147935;
PR	10-AUG-1999;	99US-0148171;

PR 11-AUG-1999; 99US-0148319.
 PR 12-AUG-1999; 99US-0148341.
 PR 13-AUG-1999; 99US-0148565.
 PR 15-AUG-1999; 99US-0148684.
 PR 16-AUG-1999; 99US-0149368.
 PR 17-AUG-1999; 99US-0149175.
 PR 18-AUG-1999; 99US-0149426.
 PR 20-AUG-1999; 99US-0149722.
 PR 20-AUG-1999; 99US-0149723.
 PR 20-AUG-1999; 99US-0149929.
 PR 23-AUG-1999; 99US-0149902.
 PR 23-AUG-1999; 99US-0149930.
 PR 25-AUG-1999; 99US-0150566.
 PR 26-AUG-1999; 99US-0150884.
 PR 27-AUG-1999; 99US-0151065.
 PR 27-AUG-1999; 99US-0151066.
 PR 27-AUG-1999; 99US-0151080.
 PR 30-AUG-1999; 99US-0151303.
 PR 31-AUG-1999; 99US-0151438.
 PR 01-SEP-1999; 99US-0151930.
 PR 07-SEP-1999; 99US-0152363.
 PR 10-SEP-1999; 99US-0153070.
 PR 13-SEP-1999; 99US-0153758.
 PR 15-SEP-1999; 99US-0154018.
 PR 16-SEP-1999; 99US-0154039.
 PR 20-SEP-1999; 99US-0154779.
 PR 22-SEP-1999; 99US-0155139.
 PR 23-SEP-1999; 99US-0155486.
 PR 24-SEP-1999; 99US-0155639.
 PR 28-SEP-1999; 99US-0156458.
 PR 28-SEP-1999; 99US-0156596.
 PR 04-OCT-1999; 99US-0157117.
 PR 05-OCT-1999; 99US-0157753.
 PR 06-OCT-1999; 99US-0157865.
 PR 07-OCT-1999; 99US-0158029.
 PR 08-OCT-1999; 99US-0158232.
 PR 12-OCT-1999; 99US-0158369.
 PR 13-OCT-1999; 99US-0159293.
 PR 13-OCT-1999; 99US-0159294.
 PR 13-OCT-1999; 99US-0159295.
 PR 14-OCT-1999; 99US-0159329.
 PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159637.
 PR 14-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159538.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 23-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161922.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.
 PR 29-OCT-1999; 99US-0162143.

Query Match 99.74%; Score 2446; DB 21; Length 485;
 Best Local Similarity 99.8%; Pred No 3; le-209;
 Matches 465; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 1 MSSPLGHQILVFLHALLIFSASRKTQLLNDNDVSSDKSAGTRWAVLVAGSNEYNY 60
 20 MSSPLGHQILVFLHALLIFSASRKTQLLNDNDVSSDKSAGTRWAVLVAGSNEYNY 79

Qy 61 RHQADICHAYOILRKGLKNDENIIVFWDDIAFSSSPRGVLIINKPGCEDVYKGVPKDY 120
 Db 80 RHQADICHAYOILRKGLKNDENIIVFWDDIAFSSSPRGVLIINKPGCEDVYKGVPKDY 139
 Qy 121 TKEAVNVQNFYNYLLNENSGVTGNGKVKVSGPNDNFIYYADHGAAPGLIAMPTGDEVMA 180
 Db 140 TKEAVNVQNFYNYLLNENSGVTGNGKVKVSGPNDNFIYYADHGAAPGLIAMPTGDEVMA 199
 Qy 181 KDFNEVLEKMKHRKRYKMKVIYYEACESGSMFEGILKKNLIYAVYTAANSKESSGVYCP 240
 Db 200 KDFNEVLEKMKHRKRYKMKVIYYEACESGSMFEGILKKNLIYAVYTAANSKESSGVYCP 259
 Qy 241 ESYPPPPSEIGTCLGDTFESISWLEDSDLHDMSKETLEQQYHVVKRRVSGSDVPETSHVCRF 300
 Db 260 ESYPPPPSEIGTCLGDTFESISWLEDSDLHDMSKETLEQQYHVVKRRVSGSDVPETSHVCRF 319
 Qy 301 GTEKMLKDYLLSSYIGRNPNENDNFTTETSESSPISNSGLVNPDPILLYLQRIKAPMGS 360
 Db 320 GTEKMLKDYLLSSYIGRNPNENDNFTTETSESSPISNSGLVNPDPILLYLQRIKAPMGS 379
 Qy 361 LESKEAQKLLDEKKNRKQIDQSITDILRLSVKQTNVNLNLTSTRTTGOPLVDDWDCPKT 420
 Db 380 LESKEAQKLLDEKKNRKQIDQSITDILRLSVKQTNVNLNLTSTRTTGOPLVDDWDCPKT 439
 Qy 421 LVNSFKNHCATVHYGLKYGALANICNMGVQVVKQVSAIEQACSM 466
 Db 440 LVNSFKNHCATVHYGLKYGALANICNMGVQVVKQVSAIEQACSM 485

RESULT 4
 AAG52574
 ID AAG52574 standard; Protein; 380 AA.
 XX AAG52574;
 AC AAG52574;
 DT 18-OCT-2000 (first entry)
 XX
 XX Arabidopsis thaliana protein fragment SEQ ID NO: 66848.
 DE Protein identification; signal transduction pathway; metabolic pathway;
 KW hybridisation assay; genetic mapping; gene expression control; promoter;
 KW termination sequence.
 XX Arabidopsis thaliana.
 OS Arabidopsis thaliana.
 PN EP1033405-A2.
 XX
 XX 06-SEP-2000.
 XX
 XX 25-FEB-2000; 2000EP-0301439.
 XX
 XX 25-FEB-1999; 99US-0121825.
 PR 05-MAR-1999; 99US-0123180.
 PR 09-MAR-1999; 99US-0123548.
 PR 23-MAR-1999; 99US-0125788.
 PR 25-MAR-1999; 99US-0126264.
 PR 29-MAR-1999; 99US-0126785.
 PR 01-APR-1999; 99US-0127462.
 PR 06-APR-1999; 99US-0128234.
 PR 08-APR-1999; 99US-0128714.
 PR 16-APR-1999; 99US-0129845.
 PR 19-APR-1999; 99US-0130077.
 PR 21-APR-1999; 99US-0130449.
 PR 23-APR-1999; 99US-0130510.
 PR 23-APR-1999; 99US-0130891.
 PR 28-APR-1999; 99US-0131449.
 PR 30-APR-1999; 99US-0132048.
 PR 30-APR-1999; 99US-0132407.
 PR 04-MAY-1999; 99US-0132484.
 PR 05-MAY-1999; 99US-0132485.
 PR 06-MAY-1999; 99US-0132486.
 PR 06-MAY-1999; 99US-0132487.

PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136592.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 24-JUN-1999; 99US-0140823.
PR 28-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.

PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 18-OCT-1999; 99US-0159638.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.

```

PR 22-OCT-1999: 99US-0160980.
PR 22-OCT-1999: 99US-0160981.
PR 22-OCT-1999: 99US-0160989.
PR 22-OCT-1999: 99US-0161404.
PR 22-OCT-1999: 99US-0161405.
PR 22-OCT-1999: 99US-0161406.
PR 22-OCT-1999: 99US-0161359.
PR 22-OCT-1999: 99US-0161360.
PR 22-OCT-1999: 99US-0161361.
PR 22-OCT-1999: 99US-0161920.
PR 22-OCT-1999: 99US-0161992.
PR 22-OCT-1999: 99US-0161993.
PR 22-OCT-1999: 99US-0162142.
PR 22-OCT-1999: 99US-0162143.
PR 22-OCT-1999: 99US-0162144.

Query Match      81.9%; Score 2009; DB 21; Length 380;
Best Local Similarity 100.0%; Pred. No. 2.1e-170;
Matches 380; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 87 MYDDIAFSSNPFGVIINKPGEDYKGVPKDYTKAEAVNVQNFYVLLGNESGVTGGNG 146
Db 1 MYDDIAFSSNPFGVIINKPGEDYKGVPKDYTKAEAVNVQNFYVLLGNESGVTGGNG 60

QY 147 KVKSGPNDNIFYIYADHAGPLIAMPTGDEVMKDFNEVLEKMKHKKYKMKVIYVEAC 206
Db 61 KVKSGPNDNIFYIYADHAGPLIAMPTGDEVMKDFNEVLEKMKHKKYKMKVIYVEAC 120

QY 207 ESGSMFEGILKKNLNIYAVTAANSKSSGVCPSYPPPEIGTCIGDTPFSISWLEDS 266
Db 121 ESGSMFEGILKKNLNIYAVTAANSKSSGVCPSYPPPEIGTCIGDTPFSISWLEDS 180

QY 267 DLHDMKETLEQQYHVVKRRVGSVDVPTSHVCRFGTEKMLKDYLSYIGRNPENDNFTFT 326
Db 181 DLHDMKETLEQQYHVVKRRVGSVDVPTSHVCRFGTEKMLKDYLSYIGRNPENDNFTFT 240

QY 327 ESFSSPISNSGLVNPRIPLLYLQRIQKAPMGSLKESKQAQKLLDEKHNKQIDOSITD 386
Db 241 ESFSSPISNSGLVNPRIPLLYLQRIQKAPMGSLKESKQAQKLLDEKHNKQIDOSITD 300

QY 387 ILRLSVKQTNVLLNSTRTTGOPLVDDWDCFKTLVNSFKNHCAGATVHYGLKYTGALANI 446
Db 301 ILRLSVKQTNVLLNSTRTTGOPLVDDWDCFKTLVNSFKNHCAGATVHYGLKYTGALANI 360

QY 447 CNMGVDVQKTVSAIEQAQCSM 466
Db 361 CNMGVDVQKTVSAIEQAQCSM 380

RESULT 5
AAG12662
XX AAG12662 standard; Protein; 380 AA.
XX
AC AAG12662;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 11862.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.

```

```

PR 25-MAR-1999: 99US-0126264.
PR 23-MAR-1999: 99US-0126785.
PR 01-APR-1999: 99US-0127462.
PR 08-APR-1999: 99US-0128234.
PR 16-APR-1999: 99US-0128714.
PR 16-APR-1999: 99US-0129845.
PR 19-APR-1999: 99US-0130077.
PR 21-APR-1999: 99US-0130449.
PR 23-APR-1999: 99US-0130510.
PR 23-APR-1999: 99US-0130891.
PR 28-APR-1999: 99US-0131449.
PR 30-APR-1999: 99US-0132048.
PR 04-MAY-1999: 99US-0132407.
PR 05-MAY-1999: 99US-0132484.
PR 06-MAY-1999: 99US-0132485.
PR 06-MAY-1999: 99US-0132486.
PR 07-MAY-1999: 99US-0132487.
PR 11-MAY-1999: 99US-0132863.
PR 14-MAY-1999: 99US-0134256.
PR 14-MAY-1999: 99US-0134218.
PR 14-MAY-1999: 99US-0134219.
PR 14-MAY-1999: 99US-0134221.
PR 14-MAY-1999: 99US-0134370.
PR 18-MAY-1999: 99US-0134768.
PR 19-MAY-1999: 99US-0134941.
PR 20-MAY-1999: 99US-0135124.
PR 21-MAY-1999: 99US-0135353.
PR 24-MAY-1999: 99US-0135629.
PR 25-MAY-1999: 99US-0136021.
PR 27-MAY-1999: 99US-0136392.
PR 28-MAY-1999: 99US-0136782.
PR 01-JUN-1999: 99US-0137222.
PR 03-JUN-1999: 99US-0137528.
PR 04-JUN-1999: 99US-0137724.
PR 07-JUN-1999: 99US-0137502.
PR 08-JUN-1999: 99US-0138094.
PR 10-JUN-1999: 99US-0138540.
PR 10-JUN-1999: 99US-0138847.
PR 14-JUN-1999: 99US-0139119.
PR 16-JUN-1999: 99US-0139452.
PR 16-JUN-1999: 99US-0139453.
PR 17-JUN-1999: 99US-0139492.
PR 18-JUN-1999: 99US-0139454.
PR 18-JUN-1999: 99US-0139455.
PR 18-JUN-1999: 99US-0139456.
PR 18-JUN-1999: 99US-0139457.
PR 18-JUN-1999: 99US-0139458.
PR 18-JUN-1999: 99US-0139459.
PR 18-JUN-1999: 99US-0139460.
PR 18-JUN-1999: 99US-0139461.
PR 18-JUN-1999: 99US-0139462.
PR 18-JUN-1999: 99US-0139463.
PR 18-JUN-1999: 99US-0139750.
PR 18-JUN-1999: 99US-0139763.
PR 21-JUN-1999: 99US-0139817.
PR 22-JUN-1999: 99US-0139899.
PR 23-JUN-1999: 99US-0140353.
PR 23-JUN-1999: 99US-0140354.
PR 24-JUN-1999: 99US-0140695.
PR 28-JUN-1999: 99US-0140823.
PR 29-JUN-1999: 99US-0140991.
PR 30-JUN-1999: 99US-0141287.
PR 01-JUL-1999: 99US-0141842.
PR 01-JUL-1999: 99US-0142154.
PR 02-JUL-1999: 99US-0142055.
PR 06-JUL-1999: 99US-0142390.
PR 08-JUL-1999: 99US-0142803.
PR 09-JUL-1999: 99US-0142920.
PR 12-JUL-1999: 99US-0142977.
PR 13-JUL-1999: 99US-0143542.
PR 14-JUL-1999: 99US-0143624.
PR 15-JUL-1999: 99US-0144005.
PR 16-JUL-1999: 99US-0144085.

```



```
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144634.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 10-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149802.
PR 23-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 26-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 10-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.

PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 81.5%; Score 2001; DB 21; Length 380;
Best Local Similarity 99.7%; Pred. No. 1.le-169;
Matches 379; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 87 MYDDIAFSSSENPRGVIIINKPDGEDVYKGVPKDYTKKAVNVQNFYNNVLLGNESGVTGGNG 146
Db 1 MYDDIAFSSSENPRGVIIINKPDGEDVYKGVPKDYTKKAVNVQNFYNNVLLGNESGVTGGNG 60
QY 147 KVKSGPNDNFIYYADHGAPGLIAMPTGDVMAKDFNEVLEKMKHKKYKMKVIYVEAC 206
Db 61 KVKSGPNDNFIYYADHGAPGLIAMPTGDVMAKDFNEVLEKMKHKKYKMKVIYVEAC 120
QY 207 EGSFMFEGILKKNLNIYAVTAANSKESWGVCYCPESYPPPPSEIGTCIGDTFSISWLEDS 266
Db 121 EGSFMFEGILKKNLNIYAVTAANSKESWGVCYCPESYPPPPSEIGTCIGDTFSISWLEDS 180
QY 267 DLHDMSKETLEQQYHVVKRRVGSVPETSHVCREGTEKMLKDYLLSSYIGRNPENDNFTFT 326
Db 181 DLHDMSKETLEQQYHVVKRRVGSVPETSHVCREGTEKMLKDYLLSSYIGRNPENDNFTFT 240
QY 327 ESFSSPISNSGLVNPRIPLLYLQRKIQKAPMGSLKESKEAQKLLDEKNHRKQIDQSTTD 386
Db 241 ESFSSPISNSGLVNPRIPLLYLQRKIQKAPMGSLKESKEAQKLLDEKNHRKQIDQSTTD 300
QY 387 ILRLSVKQTNVNLTLSTRTTGGQPLVDWDCFKTLVNSFKNHCATVHYGLKYTGALANI 446
Db 301 ILRLSVKQTNVNLTLSTRTTGGQPLVDWDCFKTLVNSFKNHCATVHYGLKYTGALANI 360
QY 447 CNMGVDVKQTVSAIEQACSM 466
Db 361 CNMGVDVKQTVSAIEQACSM 380

RESULT 6
AAG52575
ID AAG52575 standard; Protein; 295 AA.
XX
AC AAG52575;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 66849.
```



```

Db 300 QRDADLIHFEKFAPEGSPEKNAQKQVVEVMSHRMHIDNGVELIGKLLFGIEKGPVK 359
QY 401 LTPRTTGOPLVDWDCFKTLVNSFKNHCGATVHYGLKYTGALANICNGVDVKOTVSAI 460
DB 360 LDVAPGAMALVDDWDCFKTLVNTFTFYCGSLQYCKMKHRSFANICNAGINKDQADAS 419
QY 461 EQAC 464
DB 420 AQAC 423

RESULT 8
AAG31187
ID AAG31187 standard; Protein; 490 AA.
AC AAG31187;
DT 17-OCT-2000 (first entry)
DE Arabidopsis thaliana protein fragment SEQ ID NO: 37411.
XX
XX Protein identification; signal transduction pathway; metabolic pathway;
XX hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
XX Arabidopsis thaliana.
XX
XX EP1033405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123180.
XX 09-MAR-1999; 99US-0123548.
XX 23-MAR-1999; 99US-0125788.
XX 25-MAR-1999; 99US-0126264.
XX 29-MAR-1999; 99US-0126785.
XX 01-APR-1999; 99US-0127462.
XX 04-APR-1999; 99US-0128234.
XX 08-APR-1999; 99US-0128714.
XX 16-APR-1999; 99US-0128645.
XX 19-APR-1999; 99US-0130077.
XX 21-APR-1999; 99US-0130449.
XX 23-APR-1999; 99US-0130510.
XX 28-APR-1999; 99US-0130891.
XX 30-APR-1999; 99US-0131449.
XX 30-APR-1999; 99US-0132048.
XX 04-MAY-1999; 99US-0132407.
XX 05-MAY-1999; 99US-0132484.
XX 06-MAY-1999; 99US-0132485.
XX 07-MAY-1999; 99US-0132486.
XX 11-MAY-1999; 99US-0132487.
XX 14-MAY-1999; 99US-0132486.
XX 14-MAY-1999; 99US-0134218.
XX 14-MAY-1999; 99US-0134219.
XX 14-MAY-1999; 99US-0134221.
XX 14-MAY-1999; 99US-0134370.
XX 18-MAY-1999; 99US-0134768.
XX 19-MAY-1999; 99US-0134941.
XX 20-MAY-1999; 99US-0135124.
XX 21-MAY-1999; 99US-0135353.
XX 24-MAY-1999; 99US-0135629.
XX 25-MAY-1999; 99US-0136021.
XX 27-MAY-1999; 99US-0136392.
XX 28-MAY-1999; 99US-0136782.
XX 01-JUN-1999; 99US-0137222.
XX 04-JUN-1999; 99US-0137528.
XX 07-JUN-1999; 99US-0137502.
XX 08-JUN-1999; 99US-0137724.
XX 08-JUN-1999; 99US-0138094.
XX
XX 10-JUN-1999; 99US-0138540.
XX 10-JUN-1999; 99US-0138847.
XX 14-JUN-1999; 99US-0139119.
XX 16-JUN-1999; 99US-0139452.
XX 16-JUN-1999; 99US-0139453.
XX 17-JUN-1999; 99US-0139492.
XX 18-JUN-1999; 99US-0139451.
XX 18-JUN-1999; 99US-0139455.
XX 18-JUN-1999; 99US-0139456.
XX 18-JUN-1999; 99US-0139457.
XX 18-JUN-1999; 99US-0139458.
XX 18-JUN-1999; 99US-0139459.
XX 18-JUN-1999; 99US-0139460.
XX 18-JUN-1999; 99US-0139461.
XX 18-JUN-1999; 99US-0139462.
XX 18-JUN-1999; 99US-0139463.
XX 18-JUN-1999; 99US-0139750.
XX 18-JUN-1999; 99US-0139763.
XX 21-JUN-1999; 99US-0139817.
XX 22-JUN-1999; 99US-0139899.
XX 23-JUN-1999; 99US-0140353.
XX 23-JUN-1999; 99US-0140354.
XX 24-JUN-1999; 99US-0140695.
XX 28-JUN-1999; 99US-0140823.
XX 29-JUN-1999; 99US-0140991.
XX 30-JUN-1999; 99US-0141287.
XX 01-JUL-1999; 99US-0141842.
XX 01-JUL-1999; 99US-0142154.
XX 02-JUL-1999; 99US-0142055.
XX 06-JUL-1999; 99US-0142390.
XX 08-JUL-1999; 99US-0142803.
XX 09-JUL-1999; 99US-0142920.
XX 12-JUL-1999; 99US-0142977.
XX 13-JUL-1999; 99US-0143424.
XX 14-JUL-1999; 99US-0143624.
XX 15-JUL-1999; 99US-0144005.
XX 16-JUL-1999; 99US-0144085.
XX 16-JUL-1999; 99US-0144086.
XX 19-JUL-1999; 99US-0144325.
XX 19-JUL-1999; 99US-0144331.
XX 19-JUL-1999; 99US-0144332.
XX 19-JUL-1999; 99US-0144333.
XX 19-JUL-1999; 99US-0144334.
XX 19-JUL-1999; 99US-0144335.
XX 20-JUL-1999; 99US-0144352.
XX 20-JUL-1999; 99US-0144632.
XX 20-JUL-1999; 99US-0144884.
XX 21-JUL-1999; 99US-0144814.
XX 21-JUL-1999; 99US-0145086.
XX 21-JUL-1999; 99US-0145088.
XX 22-JUL-1999; 99US-0145085.
XX 22-JUL-1999; 99US-0145087.
XX 22-JUL-1999; 99US-0145089.
XX 22-JUL-1999; 99US-0145192.
XX 23-JUL-1999; 99US-0145145.
XX 23-JUL-1999; 99US-0145218.
XX 23-JUL-1999; 99US-0145224.
XX 26-JUL-1999; 99US-0145276.
XX 27-JUL-1999; 99US-0145913.
XX 27-JUL-1999; 99US-0145918.
XX 27-JUL-1999; 99US-0145919.
XX 28-JUL-1999; 99US-0145951.
XX 02-AUG-1999; 99US-0146386.
XX 02-AUG-1999; 99US-0146388.
XX 02-AUG-1999; 99US-0146389.
XX 03-AUG-1999; 99US-0147038.
XX 04-AUG-1999; 99US-0147204.
XX 04-AUG-1999; 99US-0147302.
XX 05-AUG-1999; 99US-0147192.
XX 05-AUG-1999; 99US-0147260.
XX 06-AUG-1999; 99US-0147303.
XX 06-AUG-1999; 99US-0147416.
XX 09-AUG-1999; 99US-0147493.

```

PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 50.8%; Score 1246; DB 21; Length 490;
Best Local Similarity 53.6%; Pred. No. 3.4e-102;
Matches 233; Conservative 71; Mismatches 121; Indels 10; Gaps 4;
OY 36 ESSDKSAGTRWAVLVAGSNEYNYRHQADICHAYQLRKGLKDENIIVFMYDDIAFSS 95

Db 45 ENDDSSNGSTRWAVLVAGSSGYWNYRHQADICHAYQLRKGLKEENIVFMYDDIANNY 104
OY 96 ENPRPGVIINKPGEDYVKGVPKDYTKKAVNVQNFYVLLGNESGVTTGGKGVKSGPND 155
Db 105 ENPRPGTIIINSPHGKDYQGVKDYTGDDVNVNDLFAVLGDKRTAVKGGSGKVVDSGPND 164
OY 156 NIFIYADHAGPLIAMPTGDEVMAKDFNEVLEKMKHKKYKNMVIYVEACESGSMFEGI 215
Db 165 HIFIFYSDHGGPGVLGMPSTPYLVANDLNDVKKKHALGTYKSLVFFYLEACESGSIIEGL 224
OY 216 LKKNLIYAVTAANSKSSWGVCPSPYPPPSGIGTCIGDTGTFPSISWLESDLDHMSKET 275
Db 225 LPEGLNIYATTASNAESSWGTGCPGPEPSPPEYETCLGDLYSVAMMEDSGMHLQET 284
OY 276 LEQOYHVVKRR--VGSDVPETSHVCRFGTEKMLKDYLVSSVIGRNPENDNFTETE--SFS 330
Db 285 LHQOYELVKRRTAPVGSY--GSHVMQYGDVGISKNDLJLYMGTNPANDNFTPADNSLK 342
OY 331 SPISNSGLVNPRIPLLYLQKKIQKAPMGSLKESKAKKLLDEKNHKKQIDQSITDILRL 390
Db 343 PP---SRVTNQRDADLVHFWEKYRKAPEGSARKTEAQKQVLEAMSHRLHIDNSVILVKKI 399
OY 391 SVKOTNVNLITSTRTTCQPLVDWDQCFKTLVNSFKNHCGATVHYGLKYTGALANICNMG 450
Db 400 LFGISRGPEVLNKYRSAGQPLVDWNCNCKNQVRAPERHRCGSLSQYGIKHMRSFANICNAG 459
OY 451 VDKQTVSATEOACS 465
Db 460 IQMEQMBEAAEQACT 474

RESULT 9
AAG31186
ID AAG31186 standard; Protein; 494 AA.
XX AC AAG31186;
DT 17-OCT-2000 (first entry)
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 37410.
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX OS Arabidopsis thaliana.
XX PN EP1033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-0301439.
XX PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134221.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 26-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149729.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 25-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151080.
PR 27-AUG-1999; 99US-0151303.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 16-SEP-1999; 99US-0154018.
PR 20-SEP-1999; 99US-0154039.
PR 22-SEP-1999; 99US-0154779.
PR 23-SEP-1999; 99US-0155139.
PR 24-SEP-1999; 99US-0155486.
PR 28-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 07-OCT-1999; 99US-0157865.
PR 08-OCT-1999; 99US-0158029.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159631.
PR 14-OCT-1999; 99US-0159637.
PR 18-OCT-1999; 99US-0159638.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.

XX Asparaginylendopeptidase; Canavalia ensiformis; seed;
 KW L-asparagine; primer; PCR; protein fragmentation;
 KW peptide synthesis.
 OS Canavalia ensiformis.
 XX JP05276960-A.
 XX 26-OCT-1993.
 XX 07-AUG-1992; 92JP-0231602.
 XX 07-FEB-1992; 92JP-0056023.
 XX (SYUZ/) SYUZO T.
 XX WPI; 1993-373587/47.
 XX N-PSDB; AAQ50572.
 XX New gene for encoding asparaginyl endo-peptidase - comprises 8
 XX specified DNA sequences
 XX Disclosure; Page 20-22; 35pp; Japanese.
 XX A gene encoding asparaginylendopeptidase is claimed.
 CC 8 DNA sequences are given (AAQ50559-66). The enzyme is a protease
 CC derived from a seed of Canavalia ensiformis which selectively
 CC hydrolyses C-terminus amide bond of L-asparagine residue
 CC (see AAR43033 and AAR43041).
 CC The enzyme is useful for protein fragmentation and enzymatic
 CC peptide synthesis.
 CC The primers given in AAQ50567-68, AAQ50576-77 and AAQ50583-90 were used
 CC in the isolation of the fragments given in AAQ50569-75 and AAQ50578-79,
 CC by PCR.
 XX Sequence 440 AA;

Query Match 49.5%; Score 1215; DB 14; Length 440;
 Best Local Similarity 53.3%; Pred. No. 1.7e-99;
 Matches 226; Conservative 65; Mismatches 127; Indels 6; Gaps 3;
 QY 44 GTRAVLVAGSNYYNRYHQADICHAYQILRKGLKDNIIIVFYDDIAFSSNPRPGVI 103
 DB 3 GTRAVLVAGSNYYNRYHQADICHAYQILRKGLKDNIIIVFYDDIAFSSNPRPGVI 62
 QY 104 INKPGEDVYKGVKDYTKKAVNVQNFYVLLGNESGVTGGNGKVKSGPNDNFIYAD 163
 DB 63 INKPGDDVYEGVPKDYTGDDVTADNFYAVILGNKSLTGGSGKVVNSGPDRIIFYSD 122
 QY 164 HGAPGLIAMPTGDEVMAKDFNEVLKMKRKKYKMKWYIYVEACSGSMFEGILKKNLIY 223
 DB 123 HGGPGVLCMPAGPFLYASDLTEVLKMKRKKYKMKWYIYVEACSGSMFEGILKKNLIY 182
 QY 224 AVTAANSKSSNGVYCPESYPPPESEIGTCLGDTFSISWLEDSLDHMSKETLEQQYHV 283
 DB 183 ATTASNAESSNGVYCPESYPPPESEIGTCLGDTFSISWLEDSLDHMSKETLEQQYHV 242
 QY 284 KRR-VGSDVPTSHVCRGTETKMLKDYLSYIGRNPNDNFTPE--SFSPISNSGLN 340
 DB 243 KERTNGSIYSHSHVQYGDLSLDVLFYLGTPNDNFTYVDENSLRSP--SKAIS 299
 QY 341 PRDIPLLYLQRIQAPMGSLSEAKQKLLDEKNHRKQIDQSTIDILRLSVKOTNVLN 400
 DB 300 QRVADLIHFWEKRAPEGSRTKRAAQKQFLEVMSHRMHIDNSYKILGSLFGIEKGPV 359
 QY 401 LRSTTTTQPLVDWDCFKTLVNSFKNHCATVHYGLKYTGALANICNMGVQKQYSAI 460
 DB 360 LNAVRPAGMALVDDWDCFLKNMVRTFTCYGSLSQYGMKYMRSFANICNARKNDQMDAS 419
 QY 461 EQAC 464
 DB 420 AQAC 423

RESULT 12
 AAR43040
 ID AAR43040 standard; Protein; 440 AA.
 XX AAR43040;
 XX 24-MAY-1994 (first entry)
 XX Asparaginylendopeptidase ASN.
 XX Asparaginylendopeptidase; Canavalia ensiformis; seed;
 KW L-asparagine; primer; PCR; protein fragmentation;
 KW peptide synthesis.
 XX Canavalia ensiformis.
 XX JP05276960-A.
 XX 26-OCT-1993.
 XX 07-AUG-1992; 92JP-0231602.
 XX 07-FEB-1992; 92JP-0056023.
 XX (SYUZ/) SYUZO T.
 XX WPI; 1993-373587/47.
 XX N-PSDB; AAQ50579.
 XX New gene for encoding asparaginyl endo-peptidase - comprises 8
 XX specified DNA sequences
 XX Disclosure; Page 27-29; 35pp; Japanese.
 XX A gene encoding asparaginylendopeptidase is claimed.
 CC 8 DNA sequences are given (AAQ50559-66). The enzyme is a protease
 CC derived from a seed of Canavalia ensiformis which selectively
 CC hydrolyses C-terminus amide bond of L-asparagine residue
 CC (see AAR43033 and AAR43041).
 CC The enzyme is useful for protein fragmentation and enzymatic
 CC peptide synthesis.
 CC The primers given in AAQ50567-68, AAQ50576-77 and AAQ50583-90 were used
 CC in the isolation of the fragments given in AAQ50569-75 and AAQ50578-79,
 CC by PCR.
 XX Sequence 440 AA;
 Query Match 44.0%; Score 1080; DB 14; Length 440;
 Best Local Similarity 48.3%; Pred. No. 1.8e-87;
 Matches 207; Conservative 73; Mismatches 129; Indels 20; Gaps 4;
 QY 44 GTRAVLVAGSNYYNRYHQADICHAYQILRKGLKDNIIIVFYDDIAFSSNPRPGVI 103
 DB 3 GTRAVLVAGSNYYNRYHQADICHAYQILRKGLKDNIIIVFYDDIAFSSNPRPGVI 62
 QY 104 INKPGEDVYKGVKDYTKKAVNVQNFYVLLGNESGVTGGNGKVKSGPNDNFIYAD 163
 DB 63 INKPGDDVYEGVPKDYTGDDVTADNFYAVILGNKSLTGGSGKVVNSGPDRIIFYSD 122
 QY 164 HGAPGLIAMPTGDEVMAKDFNEVLKMKRKKYKMKWYIYVEACSGSMFEGILKKNLIY 223
 DB 123 HGGPGVLCMPAGPFLYASDLTEVLKMKRKKYKMKWYIYVEACSGSMFEGILKKNLIY 182
 QY 224 AVTAANSKSSNGVYCPESYPPPESEIGTCLGDTFSISWLEDSLDHMSKETLEQQYHV 283
 DB 183 VTTASNAQNSNGVYCPESYPPPESEIGTCLGDTFSISWLEDSLDHMSKETLEQQYHV 242
 QY 284 KRRVGSVDVPT--SHVCRGTETKMLKDYLSYIGRNPNDNFTTESFSSSISNSGL- 338
 DB 243 RKRTSNNSYRFGSHVMQYGDNTNITAEKLYLHIGDPATVNF-----PPHNGNLEAK 294

PR 06-AUG-1999; 99US-0147416.
 PR 09-AUG-1999; 99US-0147493.
 PR 10-AUG-1999; 99US-0147935.
 PR 11-AUG-1999; 99US-0148171.
 PR 12-AUG-1999; 99US-0148319.
 PR 13-AUG-1999; 99US-0148341.
 PR 13-AUG-1999; 99US-0148565.
 PR 16-AUG-1999; 99US-0148684.
 PR 17-AUG-1999; 99US-0149368.
 PR 18-AUG-1999; 99US-0149175.
 PR 20-AUG-1999; 99US-0149426.
 PR 20-AUG-1999; 99US-0149722.
 PR 20-AUG-1999; 99US-0149723.
 PR 23-AUG-1999; 99US-0149923.
 PR 23-AUG-1999; 99US-0149929.
 PR 23-AUG-1999; 99US-0149930.
 PR 26-AUG-1999; 99US-0150566.
 PR 26-AUG-1999; 99US-0150884.
 PR 27-AUG-1999; 99US-0151065.
 PR 27-AUG-1999; 99US-0151066.
 PR 27-AUG-1999; 99US-0151080.
 PR 30-AUG-1999; 99US-0151303.
 PR 31-AUG-1999; 99US-0151438.
 PR 01-SEP-1999; 99US-0151930.
 PR 07-SEP-1999; 99US-0152363.
 PR 10-SEP-1999; 99US-0153070.
 PR 13-SEP-1999; 99US-0153758.
 PR 15-SEP-1999; 99US-0154018.
 PR 16-SEP-1999; 99US-0154039.
 PR 20-SEP-1999; 99US-0154779.
 PR 22-SEP-1999; 99US-0155139.
 PR 23-SEP-1999; 99US-0155486.
 PR 24-SEP-1999; 99US-0155659.
 PR 28-SEP-1999; 99US-0156458.
 PR 29-SEP-1999; 99US-0156596.
 PR 04-OCT-1999; 99US-0157117.
 PR 05-OCT-1999; 99US-0157753.
 PR 06-OCT-1999; 99US-0157865.
 PR 07-OCT-1999; 99US-0158029.
 PR 08-OCT-1999; 99US-0158232.
 PR 12-OCT-1999; 99US-0158369.
 PR 13-OCT-1999; 99US-0159293.
 PR 13-OCT-1999; 99US-0159294.
 PR 13-OCT-1999; 99US-0159295.
 PR 14-OCT-1999; 99US-0159329.
 PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159637.
 PR 18-OCT-1999; 99US-0159638.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161921.
 PR 28-OCT-1999; 99US-0161992.
 PR 29-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162144.

Query Watch 41.7%; Score 1023; DB 21; Length 395;
 Best Local Similarity 50.3%; Pred. No. 1.9e-82;
 Matches 193; Conservative 65; Mismatches 116; Indels 10; Gaps 4;

QY 87 MYDIAFESSENPRGVIINKPDGEDYKGVKPKDYTKAENVQNFYNVLLGNESGVTGGG 146
 Db 1 MYDIAANNENPRPGTIINSPHGKDYVQGVKDYTGDDVDNVDNLEAVILGDKTAVKGGSG 60
 QY 147 KVKSGPNDFIYYADHAGAPGLIAMPTGDEVMKDFNEVLEKMKKKYKKNMVTYVEAC 206
 Db 61 KVVDSGPNDFIYFSDHGGPGVGLMPTSPYLYANDLNDVLLKKAHALGTYKSLVFLYLEAC 120
 QY 207 ESGSMPEGILKKNLIYAVTAANSKSSNGVYCPESYPPRSEICTCLGDTFISINLEDS 266
 Db 121 ESGSIFEGLLPEGLNIYATASNAEESMGTYCPGEPSPPPEYETCLGDLYSVAMMEDS 180
 QY 267 DLHMSKETLEQQYHVVKRR---VGSDDVPETSHVCRFTEKMLKDYLSYIGRNPNENDF 323
 Db 181 GMHNLQETLHQOYELVKRRTPAVGYSY--GSRVMOYGDVGISKONLDLYMGTNPANDNF 238
 QY 324 TYTE--SFSPISNSGLVNPRIPLLYLQRTKOTKAPMGSLSKESKAKLLDEKNHKOID 381
 Db 239 TFDANSLKPP---SRVTNORDADLVHFEWKYKAPGSAKTEAKQVLEAMSHRLHID 295
 QY 382 QSTIDILRLSVKQTNVNLNLTSTRTTGOPLVDDWDCFKTLVNSFKNHCGATVHYGLKYTG 441
 Db 296 NSVILVCKILFGISRGPEVLNKRVSAGQPLVDDWNCNCKNQVRAFERHCGSLSOYGIKIMR 355
 QY 442 ALANICNMGVDDVKQTVSAIEQACS 465
 Db 356 SFANICNAGIQMEQMEEAASQACT 379

RESULT 14
 AAW69215
 ID AAW69215 standard; Protein; 433 AA.
 AC AAW69215;
 XX
 DT 08-OCT-1998 (first entry)
 DE Osteoclast inhibitor protein, OIP-2.
 XX Osteoclast inhibitor protein; OIP-1; OIP-2; human; osteoclastogenesis;
 KW postmenopausal osteoporosis; Paget's disease; bone metastases; therapy;
 KW rheumatoid arthritis.
 XX Homo sapiens.
 XX
 PN WO9828423-A2.
 XX
 PD 02-JUL-1998.
 XX
 PF 18-DEC-1997; 97WO-US23666.
 XX
 PR 20-DEC-1996; 96US-0772441.
 XX
 PA (TEXA) UNIV TEXAS SYSTEM.
 XX
 PI Choi S, Reddy SV, Roodman GD;
 XX
 DR WPI: 1998-377654/32.
 DR N-PSDB: AAV4686.
 XX
 PT New isolated osteoclast inhibitor protein - used to develop products
 for treating e.g osteoporosis, Paget's disease, bone metastases, or
 PT destructive rheumatoid arthritis
 XX
 PS Claim 5: Page 75-77: 96pp; English.
 XX
 CC This sequence represents the human osteoclast inhibitor protein, OIP-2,
 which can be used in the method of the invention. The method is
 CC for inhibiting osteoclastogenesis, and comprises administering to a cell
 CC an osteoclast (OCL) inhibitor protein (OIP) composition in a vehicle
 CC where the composition inhibits production of OCLs. The OIP polypeptides
 CC have activity in inhibiting release of calcium from bone and in

CC inhibiting osteoclast formation. They can be used for treating
CC e.g. postmenopausal osteoporosis, Paget's disease of bone, bone
CC metastases and destructive rheumatoid arthritis. They can also be used
CC for treating subjects at risk of developing osteoporosis. The products
CC can also be used for detection and diagnosis.

XX
SQ Sequence 433 AA;

Query Match 33.7%; Score 828; DB 19; Length 433;
Best Local Similarity 38.2%; Pred. No. 5.4e-65;
Matches 179; Conservative 80; Mismatches 157; Indels 52; Gaps 13;

QY 8 FOILVFLHALLIFSASRKQTQLLNDNDVSSDSKAKTRWAVLVAGSNEYNYNRHQADIC 67
Db 3 WKAVFLSVLALGTA-----VPIDDPEDGGKHWVIVAGSNGWNYNRHQADAC 50
QY 68 HAYQILRKGGKLDENIIIVFYDDIAFSENPRPGVIINKPDGEDVYKVPKDYTKKAVNV 127
Db 51 HAYQIIHRNGIPDEQIVVMYDDIAYSEDNPTGIVINRPNGTDVYOGVPKDYTGEDVTP 110
QY 128 ONFYNVLLGNESGVTG-GNGKVKSGPNDFIYIYADHGAPGLIAMPTGDEVMAKDFNEV 186
Db 170 IHVMYKHKYRKWVYIEACESGSMNH-LPDNINVYATTAANPRESSYACYDE----- 223
QY 247 PSEIGTCLGDTFISWLESDLDHMSKETLEQQYHVYKRVGVSDVPETSHVCRFGTEKML 306
Db 224 --KRSTVLGDWYSVNMWEDSDVEDLTRETQHLVKSHT-----NTSHVMQYGNKTIS 276
QY 307 KDYLSSYIGRNPNENDNFTTESFSS-----PISNGLVNPDRDIPLLYLQKKIQKAPMGS 361
Db 277 TMKVMQFG-----MKRKASSPVPLPVPVTHLDLTSPDPVPLTIMKRKLMT--NDL 325
QY 362 -ESKEAQKLLDEKNHRKQIDQSTIDILRLSVKQTNVNLNLTSTRTTGGQPLVDDWDCFKT 420
Db 326 EESQLTTEEIORHLDARHLIEKSVRKIVSLLAASEAEVEQLLSERA---PLTGH-SCYPE 381
QY 421 LVNSFKNHC---GATVHYGLKTYGTALANICNMGVQDVVKQVSAEQAC 464
Db 382 ALLHFRTHCFNWHSPTEYALRHLYLVNLCEKPYPLHRIKLSMDHVC 429

RESULT 15
AAB36175
ID AAB36175 standard; Protein; 433 AA.
AC AAB36175;
XX
XX 21-FEB-2001 (first entry)
DT Human FDH02 protein.
DE
DE
XX
KW Human; FDH02; protease; cytostatic; immunomodulatory; cancer;
KW degenerative disorder; antigen processing; pro-protein processing.
XX
XX Homo sapiens.
XX
XX U56140098-A.
XX
XX 31-OCT-2000.
XX
XX 30-AUG-1996; 96US-0706216.
XX
XX 30-AUG-1996; 96US-0706216.
XX
XX (SCHE) SCHERING CORP.
XX
XX Balasubramanian S, Ford J, Zurawski G, Gorman DM;
XX WPI; 2001-006328/01.

DR N-PSDB; AAC68662.
XX New nucleic acids encoding proteinases, useful in forensic assays or in
PT situ assays to detect chromosomal abnormalities, or for enhancing the
PT expression of proteases, which are useful for treating e.g. abnormal
PT proliferation

XX
PS Claim 1; Column 9-12; 35pp; English.

XX The present sequence is human FDH02 protein. The nucleotide
CC sequence encoding this protein is useful in diagnostic kits, forensic
CC assays or in an in situ assay to detect chromosomal abnormalities. The
CC protein is useful for mediating various aspects of cellular physiology
CC or development, e.g. for the conversion of pro-proteins to proteins, or
CC for proper immunological function, antigen processing and presentation.
CC The protein or its fragments is useful in treating conditions
CC associated with abnormal physiology or development, e.g. abnormal
CC proliferation in cancerous conditions, or degenerative conditions. The
CC nucleic acids and proteins are also useful for drug screening
CC techniques.

XX
SQ Sequence 433 AA;

Query Match 33.7%; Score 828; DB 22; Length 433;
Best Local Similarity 38.2%; Pred. No. 5.4e-65;
Matches 179; Conservative 80; Mismatches 157; Indels 52; Gaps 13;

QY 8 FOILVFLHALLIFSASRKQTQLLNDNDVSSDSKAKTRWAVLVAGSNEYNYNRHQADIC 67
Db 3 WKAVFLSVLALGTA-----VPIDDPEDGGKHWVIVAGSNGWNYNRHQADAC 50
QY 68 HAYQILRKGGKLDENIIIVFYDDIAFSENPRPGVIINKPDGEDVYKVPKDYTKKAVNV 127
Db 51 HAYQIIHRNGIPDEQIVVMYDDIAYSEDNPTGIVINRPNGTDVYOGVPKDYTGEDVTP 110
QY 128 ONFYNVLLGNESGVTG-GNGKVKSGPNDFIYIYADHGAPGLIAMPTGDEVMAKDFNEV 186
Db 111 ONFLAVLRGDAEAVKIGSGKVLKSGPDQHVFIYTDHGSTGLVFP-NEDLHVKDLNET 169
QY 187 LEKMHKRRKYNKYVYVEACESGSMFEGILKKNLIYAVTAANSKESWGVCYCPESYPPP 246
Db 170 IHVMYKHKYRKWVYIEACESGSMNH-LPDNINVYATTAANPRESSYACYDE----- 223
QY 247 PSEIGTCLGDTFISWLESDLDHMSKETLEQQYHVYKRVGVSDVPETSHVCRFGTEKML 306
Db 224 --KRSTVLGDWYSVNMWEDSDVEDLTRETQHLVKSHT-----NTSHVMQYGNKTIS 276
QY 307 KDYLSSYIGRNPNENDNFTTESFSS-----PISNGLVNPDRDIPLLYLQKKIQKAPMGS 361
Db 277 TMKVMQFG-----MKRKASSPVPLPVPVTHLDLTSPDPVPLTIMKRKLMT--NDL 325
QY 362 -ESKEAQKLLDEKNHRKQIDQSTIDILRLSVKQTNVNLNLTSTRTTGGQPLVDDWDCFKT 420
Db 326 EESQLTTEEIORHLDARHLIEKSVRKIVSLLAASEAEVEQLLSERA---PLTGH-SCYPE 381
QY 421 LVNSFKNHC---GATVHYGLKTYGTALANICNMGVQDVVKQVSAEQAC 464
Db 382 ALLHFRTHCFNWHSPTEYALRHLYLVNLCEKPYPLHRIKLSMDHVC 429

Search completed: May 27, 2003, 15:26:33
Job time : 46 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 27, 2003, 15:21:29 ; Search time 38 Seconds
(without alignments)
2526.788 Million cell updates/sec

Title: US-09-934-066-2

Perfect score: 2454

Sequence: 1 MSSPLGHFQILVFLHALLIF.....CNMGVDVKQTVSAIEAQCSM 466

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_21.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phase.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_rvirus.*
16: sp_bacteriap.*
17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2454	100.0	466	10	Q9LJX8
2	1266.5	51.6	483	10	Q9XFZ4
3	1254.5	51.1	483	10	Q9AUD9
4	1252	51.0	482	10	Q9XFZ5
5	1246	50.8	494	10	Q93VM0
6	1243.5	50.7	484	10	Q24325
7	1236	50.4	486	10	Q949L7
8	1228.5	50.1	501	10	Q9SSZ4
9	1219	49.7	492	10	Q9MAR6
10	1208.5	49.2	485	10	Q9ZB28
11	1204.5	49.1	485	10	Q9FER6
12	1191.5	48.6	481	10	Q9SXB2
13	1183.5	48.2	486	10	Q9SXB3
14	1182.5	48.2	485	10	Q9FER7
15	1166.5	47.5	460	10	Q9SHD0
16	1160	47.3	486	10	Q93VS7

17	1148	46.8	484	10	Q39044
18	1125	45.8	536	10	Q9SI79
19	1118.5	45.6	489	10	Q9LLO4
20	1115.5	45.5	493	10	Q24326
21	1094	44.6	380	10	Q9XGB9
22	1080	44.0	488	10	Q24539
23	1072.5	43.7	400	10	Q9XG75
24	1054	43.0	503	10	Q82102
25	1051.5	42.8	419	10	Q9XG76
26	1051	42.8	411	10	Q9ZT14
27	1046.5	42.6	496	10	Q9LLO5
28	1021.5	41.6	465	10	Q8VZY0
29	996.5	40.6	452	10	Q9LWZ3
30	843	34.4	435	11	Q89017
31	832	33.9	435	11	Q9R0J8
32	832	33.9	435	11	Q9JLN3
33	827	33.7	433	4	Q9BTY1
34	810	33.0	433	6	Q95M12
35	799	32.6	429	5	Q9NFY9
36	782.5	31.9	462	5	Q17945
37	562.5	22.9	187	5	Q9U589
38	326.5	13.3	326	5	Q46047
39	326.5	13.3	355	5	Q8T4E1
40	321.5	13.1	149	10	Q8VWT3
41	297.5	12.1	395	4	Q14822
42	297.5	12.1	396	4	Q92643
43	292.5	11.9	380	3	Q9USP5
44	287.5	11.7	332	4	Q8TC22
45	280.5	11.4	395	11	Q9CXY9

ALIGNMENTS

RESULT 1

ID Q9LJX8 PRELIMINARY; PRT; 466 AA.
AC Q9LJX8;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Vacuolar processing enzyme (proteinase).
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RA Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RX MEDLINE=20363099; PubMed=10907853;
RA Nakamura Y.;
RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.
RT Sequence features of the regions of 4,251,695 bp covered by ninety P1.
RT TAC and BAC clones.";
RL DNA Res. 7:217-221(2000).
DR EMBL; AP000383; BAB01880.1; -.
DR InterPro; IPR001096; Legumain.
DR Pfam; PF01650; Peptidase_C13; 1.
DR PRINTS; PR00776; HEMOGLOBINASE.
SQ SEQUENCE 466 AA; 52085 MW; 3192DD88972324CC CRC64;

Query Match 100.0%; Score 2454; DB 10; Length 466;
Best Local Similarity 100.0%; Pred. No. 1.7e-166;
Matches 466; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MSSPLGHFQILVFLHALLIFSAESKTKTOLLNDNVESDKSAKGTTRWAVLVAGSYNYN 60
|||||

Db 1 MSSPLGHFQILVFLHALLIFSAESRKTQLLNDNDVSSDKSAGTRWAVLVAGSNEYNY 60
Qy 61 RHQADICHAYQILRKGKLDENITVFMYYDDIAFSSSNPRPGVINKPDGDDVYKGPVKDY 120
Db 61 RHQADICHAYQILRKGKLDENITVFMYYDDIAFSSSNPRPGVINKPDGDDVYKGPVKDY 120
Qy 121 TKEAVNQVFNVLGNESGVTGNGKRVKSGPNDNFIYADHAGPGLIAMPDGEVNA 180
Db 121 TKEAVNQVFNVLGNESGVTGNGKRVKSGPNDNFIYADHAGPGLIAMPDGEVNA 180
Qy 181 KDFNEVLEKMKHKKRYKMYIYVEACSGSMFEGILKKNLIYAVTAANSKSSWGVCYCP 240
Db 181 KDFNEVLEKMKHKKRYKMYIYVEACSGSMFEGILKKNLIYAVTAANSKSSWGVCYCP 240
Qy 241 ESYPPSEITGCTGDTFSTSWLESDSLHDMKSETLEQQYHVVKRRVGVSDVPSHVCRF 300
Db 241 ESYPPSEITGCTGDTFSTSWLESDSLHDMKSETLEQQYHVVKRRVGVSDVPSHVCRF 300
Qy 301 GTEKMLKDYSSYIGRNPENDNFTFESSPISNSGLVNPRIPLLYLQRIQKAPMGS 360
Db 301 GTEKMLKDYSSYIGRNPENDNFTFESSPISNSGLVNPRIPLLYLQRIQKAPMGS 360
Qy 361 LESKEAOKLLDEKKNRQIQDSTIDILRLSVKQTNVNLNLTSTRTTGGQPLVDDWDCFKT 420
Db 361 LESKEAOKLLDEKKNRQIQDSTIDILRLSVKQTNVNLNLTSTRTTGGQPLVDDWDCFKT 420
Qy 421 LVNSFKHCGATVHYGLKYTGALANICNMGVDVKQTVSAIEQACSM 466
Db 421 LVNSFKHCGATVHYGLKYTGALANICNMGVDVKQTVSAIEQACSM 466

RESULT 2
Q9XF24 ID Q9XF24 PRELIMINARY: PRT: 483 AA.
AC Q9XF24
DT 01-NOV-1999 (T-EMBLrel. 12, Created)
DT 01-NOV-1999 (T-EMBLrel. 12, Last sequence update)
DT 01-MAR-2002 (T-EMBLrel. 20, Last annotation update)
DE Asparaginyl endopeptidase (VmpE-1)
OS Vigna mungo (Rice bean) (Black gram)
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Vigna.
OX NCBI_TaxID=3915;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-COTILEDON;
RA MEDLINE=99178794; PubMed=10080709;
RA Okamoto T., Minamikawa T.;
RT "Molecular cloning and characterization of Vigna mungo processing
RT enzyme 1 (VmpE-1), an asparaginyl endopeptidase possibly involved in
RT post-translational processing of a vacuolar cysteine endopeptidase
RT (SH-EP).";
RL Plant Mol. Biol. 39:63-73(1999).
DR EMBL; D89971; BAA76744.1; -;
DR MEROPS; C13.002; -;
DR InterPro; IPR001096; Legumain.
DR Pfam; PF01650; Peptidase_C13; 1.
DR PRINTS; PR00776; HEMOGLOBINASE.
SQ SEQUENCE 483 AA; 52982 MW; 5141734FFAC308C1 CRC64;

Query Match 51.6%; Score 1266.5; DB 10; Length 483;
Best Local Similarity 51.7%; Pred. No. 5.3e-82;
Matches 239; Conservative 75; Mismatches 139; Indels 9; Gaps 3;

Qy 11 LVFLHALLIFSAESRKTQLLNDNDVSSDKSAGTRWAVLVAGSNEYNYRHQ 63
Db 6 IIFVAVNLITLVSGRDEILRMSEASRFFQAPATDEGTRWAVLVAGSNEYNYRHQ 65
Qy 64 ADICHAYQILRKGKLDENITVFMYYDDIAFSSSNPRPGVINKPDGDDVYKGPVKDYKE 123
Db 66 SDVCHAYQILRKGKLDENITVFMYYDDIAFSSSNPRPGVINKPDGDDVYKGPVKDYKE 125

Qy 124 AVNQVFNVLGNESGVTGNGKRVKSGPNDNFIYADHAGPGLIAMPDGEVNAKDF 183
Db 124 AVNQVFNVLGNESGVTGNGKRVKSGPNDNFIYADHAGPGLIAMPDGEVNAKDF 183
Qy 184 NEVLEKMKHKKRYKMYIYVEACSGSMFEGILKKNLIYAVTAANSKSSWGVCYCPESY 243
Db 184 NEVLEKMKHKKRYKMYIYVEACSGSMFEGILKKNLIYAVTAANSKSSWGVCYCPESY 243
Qy 244 PPPSEITGCTGDTFSTSWLESDSLHDMKSETLEQQYHVVKRRVGVSDVPSHVCRFT 302
Db 244 PPPSEITGCTGDTFSTSWLESDSLHDMKSETLEQQYHVVKRRVGVSDVPSHVCRFT 302
Qy 303 EKMLKDYSSYIGRNPENDNFTFESSPISNSGLVNPRIPLLYLQRIQKAPMGSLE 362
Db 303 EKMLKDYSSYIGRNPENDNFTFESSPISNSGLVNPRIPLLYLQRIQKAPMGSLE 362
Qy 363 SKEAOKLLDEKKNRQIQDSTIDILRLSVKQTNVNLNLTSTRTTGGQPLVDDWDCFKTLV 422
Db 363 SKEAOKLLDEKKNRQIQDSTIDILRLSVKQTNVNLNLTSTRTTGGQPLVDDWDCFKTLV 422
Qy 423 NSFKHCGATVHYGLKYTGALANICNMGVDVKQTVSAIEQAC 464
Db 423 NSFKHCGATVHYGLKYTGALANICNMGVDVKQTVSAIEQAC 464

RESULT 3
Q9AUD9 ID Q9AUD9 PRELIMINARY: PRT: 483 AA.
AC Q9AUD9
DT 01-JUN-2001 (T-EMBLrel. 17, Created)
DT 01-JUN-2001 (T-EMBLrel. 17, Last sequence update)
DT 01-MAR-2002 (T-EMBLrel. 20, Last annotation update)
DE Asparaginyl endopeptidase.
GN PEL.
OS Phaseolus aureus (Mung bean) (Vigna radiata).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Vigna.
OX NCBI_TaxID=3916;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. TAINAN 5; TISSUE=COTYLEDON;
RA Yang W.H., Chen C.S.;
RT "Vigna radiata mRNA for asparaginyl endopeptidase, complete cds.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF238384; AAK15049.1; -;
DR MEROPS; C13.002; -;
DR InterPro; IPR001096; Legumain.
DR Pfam; PF01650; Peptidase_C13; 1.
DR PRINTS; PR00776; HEMOGLOBINASE.
SQ SEQUENCE 483 AA; 53305 MW; 733B7E283EFB9264 CRC64;

Query Match 51.1%; Score 1254.5; DB 10; Length 483;
Best Local Similarity 54.1%; Pred. No. 3.8e-81;
Matches 231; Conservative 71; Mismatches 124; Indels 1; Gaps 1;

Qy 38 SDKSAKGTWAVLVAGSNEYNYRHQADICHAYQILRKGKLDENITVFMYYDDIAFSSN 97
Db 41 NDDNVKGTWAVLVAGSNEYNYRHQADICHAYQILRKGKLDENITVFMYYDDIAFSSN 100
Qy 98 PRPVIINKPDGDDVYKGPVKDYKEAVNQVFNVLGNESGVTGNGKRVKSGPNDNI 157
Db 101 PRPVIINKPDGDDVYKGPVKDYKEAVNQVFNVLGNESGVTGNGKRVKSGPNDNI 160
Qy 158 FIYADHAGPGLIAMPDGEVNAKDFNEVLEKMKHKKRYKMYIYVEACSGSMFEGILK 217
Db 161 FIYADHAGPGLIAMPDGEVNAKDFNEVLEKMKHKKRYKMYIYVEACSGSMFEGILK 220
Qy 218 KNLNIYAVTAANSKSSWGVCYCPESYPPPSSEITGCTGDTFSTSWLESDSLHDMKSETLE 277
Db 221 EDINIYATTASNAEESWGVCYCPESYPPPSSEITGCTGDTFSTSWLESDSLHDMKSETLE 280
Qy 278 QOYHVVKRRVGVSDVPSHVCRFTGTEKMLKDYLVSSYIGRNPENDNFTFESSPISNSG 337

QY 96 ENRPRGVIINKPGEDVYKGVKDYKEAVNVQNFYVNLGNESVGTGNGKVKVSGPND 155
DB 109 ENRPRGTIIINSPHGKDYQGVKPKDYTDGVDNLFVAVILGDIKGGSGKGVDSGPD 168
QY 156 NIFYADHAGPLAMPTGDEVMAKDFEVLKMKHKKYKMKVYIYVEACSGSMFEG 215
DB 169 HIFYSDHSGPCVGMPTSPYLYANDLNDVLLKHALGTYKSLVFLYLEACSGSIFEG 228
QY 216 LKKNLNIYVTAANSKESWGVCYCPSPSEITGCTGDFSTSWLDFSDLDHMSKET 275
DB 229 LPEGLNIYATTASNAEESNGTYCPSESPPEVETCLDLYSVAMWEDSGHRLQDET 288
QY 276 LEOQYHVWR---VGSVPETSHVCRGTEKMLDYLSSYIGRPNENDFTFTE--SFS 330
DB 289 LHOQYELVLRRTAPGYSY--GSHVMQYGVGIGSKNDLQYMGTPANDNFTFADANSLK 346
QY 331 SPISNSGLNPRDIPILYQRTOKAPMGSLSKSAQKLLDEKHNKROIDOSTIDILRL 390
DB 347 PP---SRVTVNORDADLVHFWERYKRAPEGSARKTEAKQVLEAMSHRLHIDNSVLVSKI 403
QY 391 SVKQTVNLLTSTRTGQPLVDDWPCFKTLVNSFKNHCATVHYGLKYTGALANICNM 450
DB 404 LFGISRGPEVLNKRVSAGQPLVDDNNCLKNQVRAPERHCGSLSQYGIKMRSPFANICNAG 463
QY 451 VDVKQTVSAIEQAQS 465
DB 464 IOMEQMEEAASOACT 478

RESULT 6

ID 024325 PRELIMINARY; PRT: 484 AA.
AC 024325;
DT 01-NOV-1998 (T-EMBLrel. 08, Created)
DT 01-NOV-1998 (T-EMBLrel. 08, Last sequence update)
DT 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)
DE Vacuolar processing enzyme precursor (EC 3.4.22.-) (VPE)
DE (Legumin-like proteinase) (Lp1) (French bean)
OS Phaseolus vulgaris (kidney bean)
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC Crossids I. Fabales; Fabaceae; Papilionoideae; Phaseoleae; Phaseolus.
OX NCBI_TaxID=3885;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. MOLDAVIAN; TISSUE=COTYLEDON;
RA Senyuk V., Rotari V., Becker C., Zaharov A., Muentz K., Vaintraub I.;
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: ASPARAGINE-SPECIFIC ENDOPEPTIDASE INVOLVED IN THE
CC PROCESSING OF VACUOLAR SEED PROTEIN PRECURSORS INTO THE MATURE
CC FORMS.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C13; ALSO KNOWN AS THE
CC HEMOGLOBINASE FAMILY.
CC EMBL; 299956; CAB17078.1;
DR MEROPS; C13.002;
DR InterPro: IPR001096; Legumain.
DR Pfam; PF01650; Peptidase_C13; 1.
DR PRINTS; PR00776; HEMOGLOBINASE.
KW Signal; Hydrolase; Thiol protease.
FT SIGNAL 1 25 POTENTIAL.
FT PROPEP 26 44 POTENTIAL.
FT CHAIN 45 484 VACUOLAR PROCESSING ENZYME.
SQ SEQUENCE 484 AA; 53389 MW; 548DABEB99F7201D CRC64;

Query Match 50.7%; Score 1243.5; DB 10; Length 484;
Best Local Similarity 51.6%; Pred. No. 2,3e-80;
Matches 238; Conservative 78; Mismatches 132; Indels 13; Gaps 5;

QY 10 ILVFLHALLIFSASRKTOLLND-----NDVSSDKSAGTRWAVLVAGSNEYNYRHOA 64
DB 14 LLLFLVALV-----SAGRLVGDFLRLPDSGND-NVHGTRWALLFAGSSGYNYRHOA 67

QY 65 DICHAYOILRKGGKLDENIIVFMYDDIARFSSSENPRPGVIINKPGEDVYKGVKDYTKEA 124
DB 68 DICHAYQLLRKGGKLDENIIVFMYDDIARFSSSENPRPGVIINKPGEDVYKGVKDYTKED 127
QY 125 VNVONFVNLGNESVGTGNGKVKVSGPNDNFIYVADHAGPLAMPTGDEVMAKDFN 184
DB 128 VTAHNFYALLDGDKSLTGSGKGVNSGPNDFHIFYSDHSGPCVGMPTSPYLYASDLN 187
QY 185 EVLEKMKHKKYKMKVYIYVEACSGSMFEGILKKNLNIYVTAANSKESWGVCYCPSP 244
DB 188 EVLKKKHAAGTYKMLVFLYLEACSGSIFEGCLLPEDINVTATTSNADESNGTYCPGEDP 247
QY 245 PPSEITGCTGDFSTSWLDFSDLDHMSKETLEQQYHVWR---VGSVPETSHVCRGTE 303
DB 248 SPPEYSTEGLDLYSVAMWEDSGHRLRTETLHQQYKLVKERTISGGLYGGSHVMQYGV 307
QY 304 KMLKDYLSYIGRPNENDFTFESFSSPISNSGLNPRDIPILYQRTOKAPMGSLSK 363
DB 308 GLSKDILFHLVLTGDPANENLTFVDE-NSLWSSSKAVNQDADLVHFWMDKRAPEGS 366
QY 364 KEAKKLLDEKHNKROIDOSTIDILRLSVKQTVNLLTSTRTGQPLVDDWPCFKTLV 423
DB 367 NEARKQVLEVMSHRNHIDSVELVGLKLFEGIEKAPELLNAVRPAGSALVDDWDC 426
QY 424 SFKNHCATVHYGLKYTGALANICNMGVKQTVSAIEQAQ 464
DB 427 TFETHCGSLSQYGMKMRSPFANICNAGIRKEQMEEAASOACT 467

RESULT 7

ID 0949L7 PRELIMINARY; PRT: 486 AA.
AC 0949L7;
DT 01-DEC-2001 (T-EMBLrel. 19, Created)
DT 01-DEC-2001 (T-EMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)
DE Putative vacuolar processing enzyme.
OS Beta vulgaris.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Caryophyllidae; Caryophyllales; Chenopodiaceae; Beta.
OX NCBI_TaxID=161934;
RN [1]
RP SEQUENCE FROM N.A.
RA Kloos D.O., Oltmanns H., Dock C., Stahl D., Hehl R.;
RT "Isolation and molecular analysis of novel taproot expressed genes
RT from sugar beet."
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ309173; CAC43295.1;
DR MEROPS; C13.002;
DR InterPro: IPR001096; Legumain.
DR Pfam; PF01650; Peptidase_C13; 1.
SQ SEQUENCE 486 AA; 54113 MW; 057088049B85AB54 CRC64;

Query Match 50.4%; Score 1236; DB 10; Length 486;
Best Local Similarity 50.0%; Pred. No. 7,9e-80;
Matches 232; Conservative 78; Mismatches 142; Indels 12; Gaps 2;

QY 10 ILVFLHALLIFSASRKTOLLND-----DVSSDKSAGTRWAVLVAGSNEYNYN 60
DB 9 VLIVLCVLMSSVVDSR---LWMDNLIRWPSDHPISFESDDSDVGTWAVLIAGSSGYNY 65
QY 61 RHQADCHAYOILRKGGKLDENIIVFMYDDIARFSSSENPRPGVIINKPGEDVYKGVKDY 120
DB 66 RHQADCHAYOVLKGGKLDENIIVFMYDDIARFSSSENPRPGVIINKPGEDVYKGVKDY 125
QY 121 TKAVNVONFVNLGNESVGTGNGKVKVSGPNDNFIYVADHAGPLAMPTGDEVMA 180
DB 126 TGEDVTVNNFFAAILGNKDAITGSGKGVNSGPNDFHIFYSDHAGVGLMPTPYLYA 185
QY 181 KDFNEVLEKMKHKKYKMKVYIYVEACSGSMFEGILKKNLNIYVTAANSKESWGVCYCP 240
DB 186 DELIETLKEKHASGTYSKLVVYIEACSGSIFEGILPEGLNIYATTASNAVESNGTYCP 245


```
RESULT 12
Q9SBX2 ID Q9SBX2 PRELIMINARY; PRT; 481 AA.
AC Q9SBX2;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Legumain-like protease precursor.
GN SEE2B.
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OC Panicoideae; Andropogoneae; Zea.
OX NCBI_TaxID=4577;
RN [1]
RP SEQUENCE FROM N.A.
RA Donnison I.S., Griffiths C.M., Thomas A., Hoskin S.E., Bridges I.,
RA Thomas H.;
RT "Characterisation of See2, a senescence enhanced cDNA from maize with
RT homology to legumain.";
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ131718; CAB64544.1;
DR MEROPS; C13.002;
DR Pfam; PF01650; Peptidase_C13; 1.
DR PRINTS; PR00776; HEMOGLOBINASE.
KW Protease; Signal.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 486 LEGUMAIN-LIKE PROTEASE.
SQ SEQUENCE 481 AA; 52845 MW; 641DF43784626512 CRC64;

Query Match 48.6%; Score 1191.5; DB 10; Length 481;
Best Local Similarity 52.1%; Pred. No. 1.1e-76;
Matches 227; Conservative 71; Mismatches 127; Indels 11; Gaps 4;

QY 38 SSKSAK---GTRWAVLVAGSNYYNRYHQADICHAYQILRKGLKDENIIVFYDDIAF 93
DB 33 SDRAAADAVGTRWAVLVAGSNYYNRYHQADICHAYQIMKKGLKDENIIVFYDDIAH 92

QY 94 SSENPRGVIIINHPGGDVGAGVPKDYTKAVNVQNYNRYVLLGNESGVTGNGKVKVSGP 153
DB 93 SPENPRGVIIINHPGGDVGAGVPKDYTKAVNVQNYNRYVLLGNESGVTGNGKVKVSDSP 152

QY 154 NDNIIFYADHAGPLIAMPPTGDEVMAKDFNEVLEKMKHKKYKMKVYVPEACSGSWE 213
DB 153 NDHISVFSYDHGGVGLGMPYPLYGDDLDVLLKKAAGTYKSLVYLEACSGSIFE 212

QY 214 GILKKNLNTAVTAANSKESWGVCYCPESYPPPPSEIGTCGDTFISWLEDSLDHMSK 273
DB 213 GLLPNDINVTATASNAEESWGTCYCPESYPPPEYDTCGLDLYSVAMWEDSDFHNLRT 272

QY 274 ETLEQYHVVKRRVSGDVPET----SHVCRFGTEKMLKDYLLSSYIGRNPENDNFTTFESP 329
DB 273 ESLKQYVNLVKDRTA--VQDTFSYSGHVMQYGLNVLKHLFSYIGTNPANDNFTIEDN 330

QY 330 SSPISNSGLVNPRIPLLYLQKTKQAPMGCSLESKEAKLLDEKNHKKQIDQITDILR 389
DB 331 SLP-SFSKAVNQADLVYFWQYKRLADSSPEKNEARRELLEVMHRSVDSVELIGS 389

QY 390 LSVKQTNVLLTSTRITGQPLVDWDCFTLVNSFKNHGATVHYGLKYTGALANTCNM 449
DB 390 LFGSGDGPRLKAVRAAGEPLVDDWSCLKSVTRTFEACQCSLAQYGMKHMRSFANCN 449

QY 450 GVDKQTVSAIEQACS 465
DB 450 GILPEAVSKVAQAQACT 465

RESULT 13
Q9SBX3 ID Q9SBX3 PRELIMINARY; PRT; 486 AA.
AC Q9FER7;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Putative legumain precursor.

QY 36 ESSDKSAGKTWAVLVAGSNYYNRYHQADICHAYQILRKGLKDENIIVFYDDIAFSS 95
DB 40 DETDDGAGTRWAVLVAGSNYYNRYHQADICHAYQIMKKGLKDENIIVFYDDIAHSP 99

QY 96 ENPRGVIIINHPGGDVGAGVPKDYTKAVNVQNYNRYVLLGNESGVTGNGKVKVSGPND 155
DB 100 ENPRGVIIINHPGGDVGAGVPKDYTKAVNVQNYNRYVLLGNESGVTGNGKVKVSGPDD 159

QY 156 NIFYADHAGPLIAMPPTGDEVMAKDFNEVLEKMKHKKYKMKVYVPEACSGSMFEG 215
DB 160 HIFYSDHGGVGLGMPYPLYGDDLDVLLKKAAGTYKSLVYLEACSGSIFEGL 219

QY 216 LKKNLNTAVTAANSKESWGVCYCPESYPPPPSEIGTCGDTFISWLEDSLDHMSKET 275
DB 220 LPNDINVTATASNAEESWGTCYCPESYPPPEYDTCGLDLYSVAMWEDSDFHNLRTES 279

QY 276 LEQYHVVKRRVSGDVPET----SHVCRFGTEKMLKDYLLSSYIGRNPENDNFTTFESP 331
DB 280 LKQYKLVKDRTA--VHDTFSYSGHVMQYGLNVLKHLFSYIGTNPANDNFTIEDNSL 337

QY 332 PISNSGLVNPRIPLLYLQKTKQAPMGCSLESKEAKLLDEKNHKKQIDQITDILRLS 391
DB 338 P-SFSKAVNQADLVYFWQYKRLADSSPEKNEARRELLEVMHRSVDSVELIGSL 396

QY 392 VKQTNVLLTSTRITGQPLVDWDCFTLVNSFKNHGATVHYGLKYTGALANTCNMGV 451
DB 397 FGSGDGPRLKAVRAAGEPLVDDWSCLKSVTRTFEACQCSLAQYGMKHMRSFANCNAGI 456

QY 452 DVKQTVSAIEQACS 465
DB 457 LPEAVSKVAQAQACS 470

RESULT 14
Q9FER7 ID Q9FER7 PRELIMINARY; PRT; 486 AA.
AC Q9FER7;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Putative legumain precursor.
```



```
Db 3 WKAVFLSVALGICA-----VPIDDPEDGKHHVVIVAGSNGWNYRHQADAC 50
Qy 68 HAYOILRRKGLKDNIIIVFYDDIAFSSENPRPGVIINKPDGDDVYKGVPKDYTKAVNV 127
Db 51 HAYOIIHRNGIPDSQIVVMYDDIAYSEDNPTGIVINRPNGTDDVYOGVPKDYTGEDVTP 110
Qy 128 ONFYVNLGNESVGTG-GNGKVVKSGPNDNFIYYADHAGAPLIAMPTGDEYMAKDFNEV 186
Db 111 ONFLAVLRGDAEAVKIGSGKVLKSGPDHVFYIFTDHGSTGILVFP-NEDLHVKDLNET 169
Qy 187 LEKHHKRYKKNVYIYVEACESSMFGILKKNLIYAVTAANSKSSGVYCPESYPPP 246
Db 170 IHYMKHKRYKMYFYIYACESSMMNH-LPDNINVTATTAANPRESSYACYDE----- 223
Qy 247 PSEIGTCLGDTFISWLEDSLDHMSKETLEQQYHVYKRVGSDVPTSHVCRFGTEKML 306
Db 224 --KRSTYLGDMYSVNMWEDSDVEDLTRETLLKQYHLVKSHT-----NTSHVMOYGNKTIS 276
Qy 307 KDYLSSYIGRPNENDNFTTESFSS-----PISNSGLVNPDPVTHLDTSPDPVPLTIMKRKLMT--NDL 325
Db 277 TMKVMOFOG-----MKRKASSPVLPVTHLDTSPDPVPLTIMKRKLMT--NDL 325
Qy 362 -ESKEAQKLLDEKNHRKQIDOSTITDILRLSVKQTNVNLTLSTRTTGOPLVDDWDCFKT 420
Db 326 EESQLTETEIOHLDARHLIEKSVKIVSLAASEAEVEOLLSERA---PLTGH-SCYPE 381
Qy 421 LVNSFKNHC-----CATVHYGLKTYTGALANICNMGVDDVKOTVSAIEQAC 464
Db 382 ALLHFRTHCFNWHSPTYEYALRHLVYLVNLCCKEPPYPLHRIKLSMDHVC 429

RESULT 2
US-08-706-216-4
; Sequence 4, Application US/08706216
; Patent No. 6140098
; GENERAL INFORMATION:
; APPLICANT: Balasubramanian, Sriam
; APPLICANT: Ford, John
; APPLICANT: Gorman, Daniel M.
; APPLICANT: Zurawski, Gerard
; TITLE OF INVENTION: MAMMALIAN PROTEASES; RELATED REAGENTS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ONAX Research Institute
; STREET: 901 California Avenue
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304-1104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/08/706,216
; FILING DATE: 30-AUG-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Ching, Edwin P.
; REGISTRATION NUMBER: 34,090
; REFERENCE/DOCKET NUMBER: DX0613
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-852-9196
; TELEFAX: 415-496-1200
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 433 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-706-216-4
```

```
Query Match 33.7% Score 828; DB 4; Length 433;
Best Local Similarity 38.2% Pred No 7 5e-73;
Matches 179; Conservative 80; Mismatches 157; Indels 52; Gaps 13;

Qy 8 FOILVFLHALLIFSAESRKLTOLLNDNDVSSDSKAKTRWAVLVAGSNEYNYRHQADIC 67
Db 3 WKAVFLSVALGICA-----VPIDDPEDGKHHVVIVAGSNGWNYRHQADAC 50
Qy 68 HAYOILRRKGLKDNIIIVFYDDIAFSSENPRPGVIINKPDGDDVYKGVPKDYTKAVNV 127
Db 51 HAYOIIHRNGIPDSQIVVMYDDIAYSEDNPTGIVINRPNGTDDVYOGVPKDYTGEDVTP 110
Qy 128 ONFYVNLGNESVGTG-GNGKVVKSGPNDNFIYYADHAGAPLIAMPTGDEYMAKDFNEV 186
Db 111 ONFLAVLRGDAEAVKIGSGKVLKSGPDHVFYIFTDHGSTGILVFP-NEDLHVKDLNET 169
Qy 187 LEKHHKRYKKNVYIYVEACESSMFGILKKNLIYAVTAANSKSSGVYCPESYPPP 246
Db 170 IHYMKHKRYKMYFYIYACESSMMNH-LPDNINVTATTAANPRESSYACYDE----- 223
Qy 247 PSEIGTCLGDTFISWLEDSLDHMSKETLEQQYHVYKRVGSDVPTSHVCRFGTEKML 306
Db 224 --KRSTYLGDMYSVNMWEDSDVEDLTRETLLKQYHLVKSHT-----NTSHVMOYGNKTIS 276
Qy 307 KDYLSSYIGRPNENDNFTTESFSS-----PISNSGLVNPDPVTHLDTSPDPVPLTIMKRKLMT--NDL 325
Db 277 TMKVMOFOG-----MKRKASSPVLPVTHLDTSPDPVPLTIMKRKLMT--NDL 325
Qy 362 -ESKEAQKLLDEKNHRKQIDOSTITDILRLSVKQTNVNLTLSTRTTGOPLVDDWDCFKT 420
Db 326 EESQLTETEIOHLDARHLIEKSVKIVSLAASEAEVEOLLSERA---PLTGH-SCYPE 381
Qy 421 LVNSFKNHC-----CATVHYGLKTYTGALANICNMGVDDVKOTVSAIEQAC 464
Db 382 ALLHFRTHCFNWHSPTYEYALRHLVYLVNLCCKEPPYPLHRIKLSMDHVC 429

RESULT 3
US-08-928-613-2
; Sequence 2, Application US/08928613
; Patent No. 5840562
; GENERAL INFORMATION:
; APPLICANT: Diep, Dinh
; APPLICANT: Braxton, Scott M.
; APPLICANT: Delegeane, Angelo M.
; TITLE OF INVENTION: A NOVEL HUMAN CYSTEINE PROTEASE
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/08/928,613
; FILING DATE:
; CLASSIFICATION: 524
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/567,506
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Luther, Barbara J.
; REGISTRATION NUMBER: 33954
; REFERENCE/DOCKET NUMBER: PF-0048 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
```

```

;
; TELEFAX: 415-852-0195
;
; INFORMATION FOR SEQ.ID NO: 2:
;
; SEQUENCE CHARACTERISTICS:
;
; LENGTH: 431 amino acids
;
; TYPE: amino acid
;
; TOPOLOGY: linear
;
; MOLECULE TYPE: protein
;
US-08-928-613-2

```

Query Match 30.3%; Score 744; DB 2; Length 431;
Best Local Similarity 38.4%; Pred. NO. 1.4e-64;
Matches 166; Conservative 70; Mismatches 128; Indels 68; Gaps 14;

APPLICATION NUMBER: 08/928,613
 FILING DATE:
 APPLICATION NUMBER: 08/567,506
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: Luther, Barbara J.
 REGISTRATION NUMBER: 33954
 REFERENCE/DOCKET NUMBER: PF-0048 US
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 415-855-0555
 TELEFAX: 415-852-0195
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 431 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-09-193-524-2

Query Match 30.3%; Score 744; DB 3; Length 431;
Best Local Similarity 38.4%; Pred. NO. 1.4e-64;
Matches 166; Conservative 70; Mismatches 128; Indels 68; Gaps

; LENGTH: 619
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4248

Query Match
Best Local Similarity 21.5%; Pred No. 0.00013;
Matches 104; Conservative 73; Mismatches 191; Indels 116; Gaps 22;

QY 22 AERKTKQLLNDNDVSSDKSAGTRWAVLVAGSNYYNRRHQADCHAYQILRGGLKD- 80
Db 70 AELRK---VADNSDEDTGLGHTGHTMA--THGVPYNSHPHSHOSTSGRTFLVHGVIENY 124
QY 81 ENIIVMYDDIAPSSSNPRPGVIINKPGEDYVKGPKDYTKAEAVNQVFNVLGNESG 140
Db 125 EELKAEYLSDVTFSSSET-----DTEVIVQLV--DY-----FSRQGLATEDA 163
QY 141 VTGGNGKVVK-----SGPNDNIFIYYADHGAPGLIAMPTGDEVMKDFNEVLKEM 190
Db 164 FT-----KVKLLHGSYALGLLDDNDKDTIYVAKNKSPLLGVGVGEFNVIASDALAMLQTT 219
QY 191 HRRKK---YKNKVIYVEACSGSMFEGILKKNLIYAVYTAANSKSSMGVCPESYPPPP 247
Db 220 NOYKEITHDEIVKRDVTVEIKDLEGHIOQRDTYTAIDAADAEGVVDHYMLKEIHEQP 279
QY 248 S-----EIGFCLGDTFSISWLESD--LHDSKETLEQOYHV-----VKRVG 288
Db 280 AVMRRIQEQDEKGNLKDSEIINDVADADRIYVAAGT---SYHAGLVGKEFIEKWAG 336
QY 289 SDVPESHVCR---FGTEKMLKDYLSYIGRNP-----NDNFTFESFS 330
Db 337 --VPTVEVASEFVNMPLEKPLFIYSQSGETADSRVLTNKLKGLSLTITNVAG 394
QY 331 SPISNGLVNPRIPLLQYRKIQAPMGSLSKESKAQKKLLDE---KNRKOIDSITDI 387
Db 395 STLSREA---DHTLLHAGPIEAVASTKAYTAQIAVLSISQIVAKNHGRETD---VDL 447
QY 388 LRUSVKOTVNLNLTSTRTTGOPLVDD-----WDCFKTLVNSFKNHCGATVHYGLKY 439
Db 448 LRERAKVTTAI-----ETIIVDAPKMEQIATDFLKTTRNAF--FIGRTIDYNNVSL 495
QY 440 TGAL 443
Db 496 EGAL 499

RESULT 6
US-09-453-702B-257
; Sequence 257, Application US/09453702B
; Patent No. 6365723
; GENERAL INFORMATION:
; APPLICANT: Blattner, Frederick R.
; Perner, Nicole T.
; Plunkett, Guy
; Welch, Rod
; TITLE OF INVENTION: No. 6365723el Sequences of E. coli O157
; NUMBER OF SEQUENCES: 265
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 1 South Pinckney Street
; CITY: Madison
; STATE: WI
; COUNTRY: US
; ZIP: 53701-2113
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44mb storage
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 8.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/453,702B
; FILING DATE: 03-Dec-1999

; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/110,955
; FILING DATE: 04-DEC-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J.
; REGISTRATION NUMBER: 27396
; REFERENCE/DOCKET NUMBER: 960296.95017
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (608) 251-9166
; TELEFAX: (608) 251-9166
; INFORMATION FOR SEQ ID NO: 257:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3169 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 257:
US-09-453-702B-257

Query Match 4.7%; Score 115.5; DB 4; Length 3169;
Best Local Similarity 21.2%; Pred No. 0.24;
Matches 115; Conservative 83; Mismatches 188; Indels 157; Gaps 30;

QY 18 LIFSASERKTOLLNDN---DVESDSKSAK---GTRWAVLVAGSNE--YYNRRHQADICH 68
Db 2496 IITGNHGAKYTFUNSGSKTIDIHLDNRSDYDTDFRNIIFEHYTNEIFISDNQGGF-- 2553
QY 69 AYOILKRGGLKDNIIYFMYDDIAPSSSNPRPGVIINKPDGE-----DVKGVKPDYTK 123
Db 2554 VISILANATSEAAINIVFRKNMTSLDSS---GSLIYPSGDIYHISDIYK--MSGRKSF 2608
QY 124 AVNVQ-----NFYNVLGNESGVTGGNGKVVKSGPN--DNIFIYYADHGAPGLIAMPTG 175
Db 2609 KLAVERKPKDIDDIINVAILETSYLO-----IKKIPNDDSDYILCLDN--PNLSS--- 2656
QY 176 DEYMAKDFNEV-----LEKMHK-----RKYNKMWIYVEACSGSMFE 213
Db 2657 ---YTUNFNDLSGYISLWDNIRGSFTPEHKNVTNIAPNEKKYISLIGLCKLSFNIDVR 2713
QY 214 GILK-KNLNIYAVYTAANSKSSMGVY-----CPE-----SYPPPSIEIGTCICID 256
Db 2714 QALEVKNKNSYKIS---KFTWETYGDIWVSPEDRISHLELDGFGNYPFSP-ELDTPISD 2767
QY 257 TFSISN---LESDSLH---DMSKETLEQOYH--VVKRRVGSVDVPETSHVCRFGTEKML 306
Db 2768 SFSYLYDNFOIVDSVHIKLLHLNRETQITPHRIILKRYFIDSFAKTSITDR---EKNI 2824
QY 307 -----KDYLSYIGRN-----PNDNFTTSEFSSPISNSGLYNPRDIPLL 347
Db 2825 YPVICDSPDHTSDIYRHPFRIVLGNKTLYPEELVKFISTKEYLSNMDVINNVIVP--- 2882
QY 348 YLORKIQKAPMGSLSKESKAQKKLLDEKNHRKQIDQSITDILRLSVKQTNV-----LN--- 399
Db 2883 -----QKTKKKNKLSIVLSNLSNKNIDVILSGVMTGTSKIFHLNNSG 2923
QY 400 --LLTSTRTTGOPLVDDMPCFKTLVNSF--KNHCGATVHYGLKYTGALANICMGVDVVK 455
Db 2924 DLLLTTSKTHGGGV---VIFKDFINNMWKNYLTITVPIDNKLSNDRINITPMGIKIOE 2980
QY 456 TVS 458
Db 2981 TVS 2983

RESULT 7
US-09-367-583-2
; Sequence 2, Application US/09367583
; Patent No. 6451566
; GENERAL INFORMATION:
; APPLICANT: Itoh, No. 6451566uya
; TITLE OF INVENTION: METHOD FOR PRODUCING DIHYDROXYACETONE-3-PHOSPHATE

FILE REFERENCE: 06501/038001
CURRENT APPLICATION NUMBER: US/09/367,583
CURRENT FILING DATE: 1999-12-02
EARLIER APPLICATION NUMBER: JP 1997-352610
EARLIER FILING DATE: 1997-12-22
NUMBER OF SEQ ID NOS: 4
SOFTWARE: Patent in ver. 2.0
SEQ ID NO 2
LENGTH: 580
TYPE: PRT
ORGANISM: Schizosaccharomyces pombe
US-09-367-583-2

Query Match 4.7%; Score 114.5; DB 4; Length 580;
Best Local Similarity 19.3%; Pred. No. 0.019;
Matches 85; Conservative 72; Mismatches 150; Indels 133; Gaps 20;

QY 19 IFSASRKTQLLNDNDVSSOKSAGTRWALVAGSNEYNYRHQADICH---AQVILRK 75
DB 77 IFASPSKOIYTGKQVSE-----LVICKN-----YTGDIHFHGMALKEKORT 122
QY 76 GGLKDENIIVFYDDIAF-----SSENPRPG-----VIINKPDGEDYKGVPKDYTKAAYN- 126
DB 123 AGKKAEE--LIAVADDSVGRKSKGVGRGLSGTVLVHVKIAGAAAARGLP-----LEAVTT 176
QY 127 -----VQNFYNVLNKGESVGTGGNGKVKSG--PNDNIFIYYADHGAG--LIAAMPTGD 176
DB 177 IAKAAIDNLVSGIASLHVHPGHEPIAKEDMKHDEMELGMGIHNEPCCKRISIPSID 236
QY 177 EVMKADFNEVLEKMKRKKYKMKWIVVEACESGSMFEGILKKNLNIYAVTAANSKSSWG 236
DB 237 DLIAOMLKOMLDQSDKRAYVKI-----EG----- 261
QY 237 VYCPSEYPPPESEICTGLDFTFSISWLESDSLHDSKMTLEQOYHVKKRRVSGDVPETSH 296
DB 262 -----DDEVLLMNNLGLSLHESAIHKVKEALAKYKINPVRFAG-PFPTIS 310
QY 297 VCRFG-----TEKMLKDYLSYVIGNRPNENDNTFTTESFSSPSISNG--LVNPRDIPL 346
DB 311 LNLGLFGITLRTDRV-----KVEGEYSVLVDLIDQPVEAIGWPLCQPSDLK- 358
QY 347 LYLQRIKQKAPMGSLESKEAOKKL-----LDEKNHRKOIDOSITDILRLSVKQTNVNL 401
DB 359 -----SKNIGNVSIIEGQKDVSPVTDREKVRQAIVNSMENLIRAEPKITKF---- 407
QY 402 TSTRITGQPLVDWDCFKTL 421
DB 408 -----DTMAGDGCCTTL 420

RESULT 8
US-09-357-251-37
Sequence 37, Application US/09357251
Patent No. 6271441
GENERAL INFORMATION:
APPLICANT: Falco, S. Carl
APPLICANT: Famodu, Layo O.
APPLICANT: Orozco, Buddy
APPLICANT: Schwaber, James S.
TITLE OF INVENTION: Plant Aminoacyl-tRNA Synthetase
FILE REFERENCE: BB-1193
CURRENT APPLICATION NUMBER: US/09/357,251
CURRENT FILING DATE: 1999-07-20
EARLIER APPLICATION NUMBER: 60/093,530
EARLIER FILING DATE: July 21, 1998
NUMBER OF SEQ ID NOS: 37
SOFTWARE: Microsoft Office 97
SEQ ID NO 37
LENGTH: 1440
TYPE: PRT
ORGANISM: Homo sapiens
US-09-357-251-37

Query Match 4.6%; Score 112.5; DB 4; Length 1440;
Best Local Similarity 22.8%; Pred. No. 0.13;
Matches 87; Conservative 56; Mismatches 140; Indels 99; Gaps 20;

QY 51 VAGSNEY------NYRHQ-----ADICHAYQILRKGLKDNENIIVFYDDIAFSSNP-- 98
DB 562 VLKDEDFKQYVKNKSKHEELMLGDP- --LKDLKKGDIQLQRRGFFICDQPYEPVSPYS 619
QY 99 ---RPGVLIINKPDGEDYKGVPKDYTKAANNVQNFYNVLLGNESVGTGGNGKVVKSGPND 155
DB 620 CKEAPCVLIYIPDGH- --KEMPTSGSKETKVE-----ATKNET-----SAP-- 659
QY 156 NIFTIYADHGAGLIAMPTGDVMAKDFNEV-----LKMHRKKYKMKWIVVEACESG 210
DB 660 -----FKERPTPSLNNCTTSDSLVLYNRAVQGDVVRELKAKKAPKEDV----- 705
QY 211 MFEGLTKKNLNIYAVTAANSKSSWGVCPSYPPPPSEIGTCLDFTFSISWLESDSLHD 270
DB 706 --DAAVKQLLSL-----KAYEKTQCYKPGN---PPAEIGNISSNSASLESKSLYD 756
QY 271 MSKETLEQOYHVKKRRVSGDVPET-----SHVCRFG-----TEKMLKDYLSYVIGNRPN 321
DB 757 ---EVAAGGEVVRKLKAEKSPKAKINEAVECLLSLKAQYKKTKEYIP---GQPLSO 809
QY 322 NPTFTESFSSPSISNS---GLVNPRIPLLYLQRIKQKAPMGSLESKEAOKKLDEKNHRK 378
DB 810 S-----SDSSPTRNSEPAGLETP-----EAKVLFQKVAS--QGEVVRKLKTEKAPKD 854
QY 379 QIDQOSITDILRLSVKQTNVNL 400
DB 855 QVDIAVQVELLQLKAQYKSLIGV 876

RESULT 9
US-09-443-184-48
Sequence 48, Application US/09443184A
Patent No. 6372431
GENERAL INFORMATION:
APPLICANT: Cunningham, Mary Jane
APPLICANT: Zweiger, Gary
APPLICANT: Kaser, Matthew R.
APPLICANT: Panzer, Scott
APPLICANT: Selhammer, Jeffrey J.
APPLICANT: Yue, Henry
APPLICANT: Baughn, Mariah
APPLICANT: Azimzai, Valda
APPLICANT: Lal, Preeti
TITLE OF INVENTION: MAMMALIAN TOXICOLOGICAL RESPONSE MARKERS
FILE REFERENCE: PC-0007 US
CURRENT APPLICATION NUMBER: US/09/443,184A
CURRENT FILING DATE: 1999-11-19
NUMBER OF SEQ ID NOS: 138
SOFTWARE: PERL Program
SEQ ID NO 48
LENGTH: 1512
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc.feature
OTHER INFORMATION: Incyte ID No. 6372431 2302721CDI
US-09-443-184-48

Query Match 4.6%; Score 112.5; DB 4; Length 1512;
Best Local Similarity 22.8%; Pred. No. 0.14;
Matches 87; Conservative 56; Mismatches 140; Indels 99; Gaps 20;

QY 51 VAGSNEY------NYRHQ-----ADICHAYQILRKGLKDNENIIVFYDDIAFSSNP-- 98
DB 634 VLKDEDFKQYVKNKSKHEELMLGDP- --LKDLKKGDIQLQRRGFFICDQPYEPVSPYS 691
QY 99 ---RPGVLIINKPDGEDYKGVPKDYTKAANNVQNFYNVLLGNESVGTGGNGKVVKSGPND 155
DB 692 CKEAPCVLIYIPDGH- --KEMPTSGSKETKVE-----ATKNET-----SAP-- 731

QY 156 NIFIIYDHGAPLIAMPTGDEVNAKDFNEV-----LEKMKRKKYKMKVIYVEACESGS 210
Db 732 -----FKERPTPLNNCTTSEDLSVLNRYAVOGDVVRLEKAKKPKEDV----- 777
QY 211 MFEGLKKNLIYAVTRANSKESWGVCPSYPPPPSEIGTCLGDTFISWLESDLDH 270
Db 778 --DAVAKOLLSL-----KAYEKEGTQYKFGN---PRAEIGONISSNSASIIESKSLYD 828
QY 271 MSKETLEQQYHVVRKRGSDVPET-----SHVCRFG-----TEKMLDYLSYIGRNPEND 321
Db 829 -----EVAAGGEVVRKLRKAKSPKAKINEAVECLLSLKAQYKKEKTGKEYIP---GQPPLSQ 881
QY 322 NTFTESESPISNS---GLVNPRIPIPLLQRLQRIQKAPMGSLKESKRAQKLLDEKHNHRK 378
Db 882 S-----SDSSSTRNSEPAGLETP-----EAKVLFDKVAS--QGEVVRKLRKTERAKPD 926
QY 379 QIDQSITDILRSVKQTNVNLN 400
Db 927 OVDIABOELLQKQYKSLIGV 948

RESULT 10
US-08-471-033-50
; Sequence 50, Application US/08471033
; Patent No. 5770696
; GENERAL INFORMATION:
; APPLICANT: Warren, Gregory W
; APPLICANT: Koziel, Michael G
; APPLICANT: Mullins, Martha A
; APPLICANT: Nye, Gordon J
; APPLICANT: Carr, Brian
; APPLICANT: Desai, Nalini M
; APPLICANT: Kostichka, N. Kristy
; APPLICANT: Duck, Nicholas B
; APPLICANT: Estruch, Juan J
; TITLE OF INVENTION: No. 5770696el Pesticidal Proteins and Strains
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; FILING DATE: US/08/471,033
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/314,594
; FILING DATE: 09-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/218,018
; FILING DATE: 23-MAR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/037,057
; FILING DATE: 25-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Pace, Gary M.
; REGISTRATION NUMBER: P-40,403
; REFERENCE/DOCKET NUMBER: CGC 1695/CIP3/DIV7 - SOLv3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8582
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 50:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1338 amino acids

; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-471-033-50

Query Match 4.28; Score 103; DB 1: Length 1338;
Best Local Similarity 20.88; Pred. No. 0.99; 144; Indels 114; Caps 19;
Matches 83; Conservative

QY 81 ENITVFMV---DDIAFSENPRPGVIINKPDGEDVYGVKDYTKYKAVNV---ONFYNV 133
Db 343 ENITVYRCGMPFQYQISDPLSL-----KDFEQFLNTIKEDKGYMST 387
QY 134 LLGNSGVTGGKGVKSPNDNIFIIYADHCAFLIAMPPTGDEVMAKDFNEVLEKMKHR 193
Db 388 SLSSERLAAGSRKI-----ILRLQVPKSGTCAVLSAIGGFASEK---EIL--LDKD 434
QY 194 KKYKMWIYVEACESGSMFEGILK--KNLNIYAVTAANSKESWGVCPSYPPPPSEIG 251
Db 435 SKYH-----IDKVTEVILKGVRYVVDATLLTNSRGPSTPTPSPSTPTPTSDIG 484
QY 252 TCLGDTFESIWLESDLDHMSKETLEQQYHVVRKRGSDVPETSHVCRFG----- 301
Db 485 STM-KTNOISITQKNOQKEMDRKGLLYGYFKGK-----DFSNLTMFAPTRDSTLIYD 535
QY 302 -----TEKMLDYLS-SYIG--RNPENDNPTFT-----ESFSSSPISNSG----- 337
Db 536 QQTANKLLDKQOEYQSIQIRWIGLIQSKETGDTFNLSEDEQAIIIEINKKLIISNGKQKV 595
QY 338 --LVNPRDIPIILQRIQKAPMGSLKESKRAQKLLDEKHNHRKIDOSITDILR---LSV 392
Db 596 VHLEKGLVPIKIEYQSDTKFNIDSKTEKLFKIDSONOPOQVQO---DELARNPEFNK 652
QY 393 KQT-----NVNLNLT-----STRTTGQPLVDDWD 416
Db 653 KESQFLAKPSKINLFTQOMKREIDEDTDTGDSIPDLWE 692

RESULT 11
US-08-471-044-50
; Sequence 50, Application US/08471044
; Patent No. 5840868
; GENERAL INFORMATION:
; APPLICANT: Warren, Gregory W
; APPLICANT: Koziel, Michael G
; APPLICANT: Mullins, Martha A
; APPLICANT: Nye, Gordon J
; APPLICANT: Carr, Brian
; APPLICANT: Desai, Nalini M
; APPLICANT: Kostichka, N. Kristy
; APPLICANT: Duck, Nicholas B
; APPLICANT: Estruch, Juan J
; TITLE OF INVENTION: No. 5840868el Pesticidal Proteins and Strains
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; FILING DATE: US/08/471,044
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/463,483
; FILING DATE: 05-JUN-1995

FILING DATE: 25-MAR-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: Spruill, W. Murray
 REGISTRATION NUMBER: 32,943
 REFERENCE/DOCKET NUMBER: CGC 1695/CIP3
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 919-541-8615
 TELEFAX: 919-541-8689
 INFORMATION FOR SEQ ID NO: 50:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1338 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-463-483A-50

Query Match	4.2%;	Score 103;	DB 2;	Length 1338;
Best Local Similarity	20.8%;	Pred. No. 0.99;		
Matches 83;	Conservative 59;	Mismatches 144;	Indels 114;	Gaps
Qy	81	ENTIVEMV---	DDIAFSSSENPRGVLIINKPDGEDVYKGPDKYTKAENVV---	QNFYNV 1333
		:	:	:
Db	343	ENTIVTRWCGMPEFGVIGDPLPSL-----	KOFFEQFLNTIKEDKGYMST	3873
		:	:	:
Qy	134	LLGNESGVTVGGNKVYKSGPNNDIFITYADHGAPGLIAMPTGDEVNAKDFNEVLEKMKHR	1933	
		:	:	:
Db	388	SLSSERLAAFGSRKI-----	ILRLQVPKSGTGYLSAIGGFASEK---	LDKD 434
		:	:	:
Qy	194	KYNNKMVTVVEACSSGMFEGLK--	KNLNIYAVTAANSKSSNGVYCPESYPPPPSEIG	2511
		:	:	:
Db	435	SKYH-----	IDKVEVIKGVKRYVDATLLTNSRGSPSTPTSPSTPTSDIG	4841
		:	:	:

485	Db	STW-KTNQISTTQKNOOKEMDRKGLLGYFKKG-----DFSNIUMFAPTRDSTLIYD	535	
490	Db	-----TEKMLKDYLX-SYIG--RNPENDNFTET-----ESFSPISNSG----	337	
500	QY	QQTANKLLDKKQOQOYSIRWIGLIQSKETGDTFNLSDEQAIETNGKIIISNKGKEQV	595	
510	Db	--LVNPRDIPLLYLQRKIKQKAPMGSLSEAKOKLLDEKNHRKQIDQSITDILR---	392	
520	QY	VHLEKGLKVPKIKETYSQDTKFNIDSKTFKELKLFKIDSONQPOQVOO---	DELRNPEFNK	652
530	QY	KOT-----NVLNLLT-----STRTTGQPLVDDWD	416	
540	Db	-----KQF-----KQF-----KQF-----KQF-----KQF-----		
550	QY	KQF-----KQF-----KQF-----KQF-----KQF-----		
560	Db	KQF-----KQF-----KQF-----KQF-----KQF-----		
570	QY	KQF-----KQF-----KQF-----KQF-----KQF-----		
580	Db	KQF-----KQF-----KQF-----KQF-----KQF-----		
590	QY	KQF-----KQF-----KQF-----KQF-----KQF-----		
600	Db	KQF-----KQF-----KQF-----KQF-----KQF-----		
610	QY	KQF-----KQF-----KQF-----KQF-----KQF-----		
620	Db	KQF-----KQF-----KQF-----KQF-----KQF-----		
630	QY	KQF-----KQF-----KQF-----KQF-----KQF-----		
640	Db	KQF-----KQF-----KQF-----KQF-----KQF-----		
650	QY	KQF-----KQF-----KQF-----KQF-----KQF-----		
660	Db	KQF-----KQF-----KQF-----KQF-----KQF-----		
670	QY	KQF-----KQF-----KQF-----KQF-----KQF-----		
680	Db	KQF-----KQF-----KQF-----KQF-----KQF-----		
690	QY	KQF-----KQF-----KQF-----KQF-----KQF-----		
700	Db	KQF-----KQF-----KQF-----KQF-----KQF-----		
710	QY	KQF-----KQF-----KQF-----KQF-----KQF-----		
720	Db	KQF-----KQF-----KQF-----KQF-----KQF-----		
730	QY	KQF-----KQF-----KQF-----KQF-----KQF-----		
740	Db	KQF-----KQF-----KQF-----KQF-----KQF-----		
750	QY	KQF-----KQF-----KQF-----KQF-----KQF-----		
760	Db	KQF-----KQF-----KQF-----KQF-----KQF-----		
770	QY	KQF-----KQF-----KQF-----KQF-----KQF-----		
780	Db	KQF-----KQF-----KQF-----KQF-----KQF-----		
790	QY	KQF-----KQF-----KQF-----KQF-----KQF-----		
800	Db	KQF-----KQF-----KQF-----KQF-----KQF-----		
810	QY	KQF-----KQF-----KQF-----KQF-----KQF-----		
820	Db	KQF-----KQF-----KQF-----KQF-----KQF-----		
830	QY	KQF-----KQF-----KQF-----KQF-----KQF-----		
840	Db	KQF-----KQF-----KQF-----KQF-----KQF-----		
850	QY	KQF-----KQF-----KQF-----KQF-----KQF-----		
860	Db	KQF-----KQF-----KQF-----KQF-----KQF-----		
870	QY	KQF-----KQF-----KQF-----KQF-----KQF-----		
880	Db	KQF-----KQF-----KQF-----KQF-----KQF-----		
890	QY	KQF-----KQF-----KQF-----KQF-----KQF-----		
900	Db	KQF-----KQF-----KQF-----KQF-----KQF-----		
910	QY	KQF-----KQF-----KQF-----KQF-----KQF-----		
920	Db	KQF-----KQF-----KQF-----KQF-----KQF-----		
930	QY	KQF-----KQF-----KQF-----KQF-----KQF-----		
940	Db	KQF-----KQF-----KQF-----KQF-----KQF-----		
950	QY	KQF-----KQF-----KQF-----KQF-----KQF-----		
960	Db	KQF-----KQF-----KQF-----KQF-----KQF-----		
970	QY	KQF-----KQF-----KQF-----KQF-----KQF-----		
980	Db	KQF-----KQF-----KQF-----KQF-----KQF-----		
990	QY	KQF-----KQF-----KQF-----KQF-----KQF-----		
1000	Db	KQF-----KQF-----KQF-----KQF-----KQF-----		

```
RESULT 13
US-08-471-046A-50
: Sequence 50, Application US/08471046A
: Patent No. 5866326
: GENERAL INFORMATION:
: APPLICANT: Warren, Gregory W
: APPLICANT: Koziel, Michael G
: APPLICANT: Mullins, Martha A
: APPLICANT: Nye, Gordon J
: APPLICANT: Carr, Brian
: APPLICANT: Desai, Nalini M
: APPLICANT: Kostichka, N. Kristy
: APPLICANT: Duck, Nicholas B
: APPLICANT: Estruch, Juan J
: TITLE OF INVENTION: Method For Isolating Vegetative Insecticidal
: TITLE OF INVENTION: Protein Genes
: NUMBER OF SEQUENCES: 50
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: No. 5866326artis Corporation
: STREET: 3054 Cornwalis Road
: CITY: Research Triangle Park
: STATE: NC
: COUNTRY: USA
: ZIP: 27709
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30B
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/471,046A
: FILING DATE: 06-JUN-1995
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/463,483
: FILING DATE: 05-JUN-1995
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/314,594
: FILING DATE: 09-SEP-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/218,018
: FILING DATE: 23-MAR-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/037,057
: FILING DATE: 25-MAR-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: Meigs, J. Timothy
: REGISTRATION NUMBER: 38,241
: REFERENCE/DOCKET NUMBER: CGC1695/CIP3/DIV8 - SOLV4
: TELEPHONE: 919-541-8587
: TELEFAX: 919-541-8689
: INFORMATION FOR SEQ ID NO: 50:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1338 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
US-08-471-046A-50

Query Match      4.2%  Score 103; DB 2; Length 1338;
Best Local Similarity 20.8%; Pred. No. 0.99;
Matches      83;  Conservative 59; Mismatches 144; Indels 114; Gaps 19;

Qy  81 ENIVFYVY---DDIAFSENPRGVINKPDGEDVYKGVKDYTKKAVNV---QNFYVW 133
Db  343 ENITVYRWCMGPEGYQISDPLPSL-----KDFEQLNTIKEDKGYMST 387
Qy  134 LLGNESVGTGGNGKVKVSGPNDFIYADHGAPGLIAMPTGDVMAKDFNEVLEKMKHR 193
Db  388 SLSSRLAAFGSRKI-----ILRLQVPGKSTGAYLSAIGGFASEK---EIL--LDKD 434
Qy  194 KYNKMYIYVEACSGSMFEGILK--KNLNIYAVTAANSKSSWGVCYCPSPPPSEIG 251

Db  435 SKYH-----IDKYTEVLIKCVKRYVVDATLLTNSRCGPSTPTPTSPSDIG 484
Qy  252 TCLGDTFFISWLESDSLHDMKSETLEQQYHVVKRVSQSDVPETSHVCRFG----- 301
Db  485 STM-KTNOISTTKNQKQKEMDKKGLLYFYFKG-----DFSNLTMFAPTRDSTLIYD 535
Qy  302 -----TEKMLKDYLS-SYIG--RNPNQDNFTT-----ESFSSPISNSG----- 337
Db  536 QQTANKLLDKKQOEYSIRWIGLIQSKETGDTFTNLSEDEQAIIENGKIISNKKREKQV 595
Qy  338 --LVNPRDIPILYLRKTIQKAPMGSLSEKQAQKLLDQKHNHRKIDQSIITDILR---LSV 392
Db  596 VHLKGLKVLPIKIEYQSDTKFNIDSKTEKELKLFKIDSQNQPOQVOQ---DELKNPEFNK 652
Qy  393 KQT-----NVNLNLT-----STRTTGQPLVDDWD 416
Db  653 KESQEFLLAKPSKINLFTQOMKREIDEDTDTGDSIPDLWE 692

RESULT 14
US-08-470-566B-50
: Sequence 50, Application US/08470566B
: Patent No. 5872212
: GENERAL INFORMATION:
: APPLICANT: Warren, Gregory W
: APPLICANT: Koziel, Michael G
: APPLICANT: Mullins, Martha A
: APPLICANT: Nye, Gordon J
: APPLICANT: Carr, Brian
: APPLICANT: Desai, Nalini M
: APPLICANT: Kostichka, N. Kristy
: APPLICANT: Duck, Nicholas B
: APPLICANT: Estruch, Juan J
: TITLE OF INVENTION: NO. 5872212artis Corporation
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: NO. 5872212artis Corporation
: STREET: 3054 Cornwalis Road
: CITY: Research Triangle Park
: STATE: NC
: COUNTRY: USA
: ZIP: 27709
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30B
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/470,566B
: FILING DATE: 06-JUN-1995
: CLASSIFICATION: 530
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/463,483
: FILING DATE: 05-JUN-1995
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/314,594
: FILING DATE: 09-SEP-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/218,018
: FILING DATE: 23-MAR-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/037,057
: FILING DATE: 25-MAR-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: Meigs, J. Timothy
: REGISTRATION NUMBER: 38,241
: REFERENCE/DOCKET NUMBER: CGC1695/CIP3/DIV4 - SOLV4
: TELEPHONE: 919-541-8587
: TELEFAX: 919-541-8689
: INFORMATION FOR SEQ ID NO: 50:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1338 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
US-08-470-566B-50
```


THIS PAGE BLANK (USPTO)

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 27, 2003, 15:27:45 ; Search time 57 Seconds
(without alignments)
810.780 Million cell updates/sec

Title: US-09-934-066-2

Perfect score: 2454

Sequence: 1 MSSPLGHFQILVFLHALLIF.....CNMGVDVKQTVSAIEQAQCSM 466

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 375593 seqs, 99172665 residues

Total number of hits satisfying chosen parameters: 375593

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published_Applications_AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	2454	100.0	466	10	US-09-934-066-2
2	744	30.3	431	9	US-09-967-796-2
3	487	19.8	319	9	US-10-043-487-298
4	115.5	4.7	3169	9	US-10-114-170-257
5	114.5	4.7	580	9	US-10-202-211-2
6	107	4.4	966	9	US-09-738-626-3847
7	103	4.2	568	10	US-09-815-242-5140
8	102	4.2	1034	9	US-09-984-130-43
9	102	4.2	1189	9	US-09-984-130-35
10	100.5	4.1	2492	10	US-09-991-258-3
11	100	4.1	828	10	US-09-816-685-4
12	99	4.0	707	9	US-10-125-540-313
13	99	4.0	707	10	US-09-764-870-313
14	98.5	4.0	969	8	US-08-945-749-3
15	98.5	4.0	26926	9	US-09-759-508B-2
16	97	4.0	503	10	US-09-752-385-8
17	96.5	3.9	452	9	US-09-905-291A-255
18	96.5	3.9	452	9	US-09-902-853-255
19	96.5	3.9	452	9	US-09-907-824-255

ALIGNMENTS

RESULT 1

US-09-934-066-2
; Sequence 2, Application US/09934066
; Patent No. US20020108149A1
; GENERAL INFORMATION:
; APPLICANT: Gruis, Darren B.
; APPLICANT: Jung, Rudolf
; TITLE OF INVENTION: Methods of Increasing Polypeptide Accumulation in Plants
; FILE REFERENCE: 35718/237251
; CURRENT APPLICATION NUMBER: US/09/934,066
; PRIOR FILING DATE: 2001-08-21
; PRIOR APPLICATION NUMBER: US 60/226,804
; PRIOR FILING DATE: 2000-08-21
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 466
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-934-066-2

Query Match 100.0%; Score 2454; DB 10; Length 466;
Best Local Similarity 100.0%; Pred. No. 2.2e-202;
Matches 466; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MSSPLGHFQILVFLHALLIFSAESRKTOLLNDNDVSSDKSAKGTWAVLVAGSNEYNY	60
Db	1	MSSPLGHFQILVFLHALLIFSAESRKTOLLNDNDVSSDKSAKGTWAVLVAGSNEYNY	60
Qy	61	RHOADICHAYQILRKGLKDKENIIVFMYDDIAFSENPRGVIINKPDGEDVYKGPDKY	120
Db	61	RHOADICHAYQILRKGLKDKENIIVFMYDDIAFSENPRGVIINKPDGEDVYKGPDKY	120
Qy	121	TKEAVNVQNFYVLLGNESGVTGNGKVKSGPNNDTIFYYADHGAFLIAMPDGDEYMA	180
Db	121	TKEAVNVQNFYVLLGNESGVTGNGKVKSGPNNDTIFYYADHGAFLIAMPDGDEYMA	180
Qy	181	KDFNEVLEKWHKRRKKNMVIYVEACSGSMFEGILKKNLNIIYAVTANSKSSNGVYCP	240
Db	181	KDFNEVLEKWHKRRKKNMVIYVEACSGSMFEGILKKNLNIIYAVTANSKSSNGVYCP	240

Qy 241 ESYPPPEISGTCGLDPTFSISWLESDSLHDMKSETLEQQYHVVKRRVGSVDVPTSHVCRF 300
Db 241 ESYPPPEISGTCGLDPTFSISWLESDSLHDMKSETLEQQYHVVKRRVGSVDVPTSHVCRF 300
Qy 301 GTEKMLKDYLSYIGRNPENDNFTTESFSSPISNSGLVNPDRIPILLYLQRIKQAPMG 360
Db 301 GTEKMLKDYLSYIGRNPENDNFTTESFSSPISNSGLVNPDRIPILLYLQRIKQAPMG 360
Qy 361 LESKEAQKLLDEKHNKQIQDOSITDILRLSVKQTNVNLNLTSTRTTGOPLVDDMDCFKT 420
Db 361 LESKEAQKLLDEKHNKQIQDOSITDILRLSVKQTNVNLNLTSTRTTGOPLVDDMDCFKT 420
Qy 421 LVNSFKHCGATVHYGLKYTGALANICNMGVDVKQTVSAIEQACSM 466
Db 421 LVNSFKHCGATVHYGLKYTGALANICNMGVDVKQTVSAIEQACSM 466

RESULT 2

US-09-967-796-2
; Sequence 2, Application US/09967796
; Patent No. US2002015535A1
; GENERAL INFORMATION:
; APPLICANT: Diep, Dinh
; Braxton, Scott M.
; Delegeane, Angelo M.
; TITLE OF INVENTION: A NOVEL HUMAN CYSTEINE PROTEASE
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/967,796
; FILING DATE: 28-Sep-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/449,422
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Lucher, Barbara J.
; REGISTRATION NUMBER: 33954
; REFERENCE/DOCKET NUMBER: PF-0048 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-852-0195
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 431 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:

Query Match 30.3%; Score 744; DB 9; Length 431;
Best Local Similarity 38.4%; Pred. No. 1.3e-55;
Matches 166; Conservative 70; Mismatches 128; Indels 68; Gaps 14;
Qy 8 FOILVFLHALLIFSAESRKTOLLNDNVSSDKSAGKTRWVLVAGSNYYNRHQADIC 67
Db 3 WKVVVFLVAGLIGCA-----VPTDDPEDGKKHVVIVAGSNGWYNRHQADAC 50
Qy 68 HAYQILKKGKLDKDIIVFMYDDIAFSSNPFGVIINPKDGEDVYKGVPKDYTKAEAVNV 127
Db 51 HAYQIFHRNGIPAEQIVIMYDDIAFSSNPFGVIINPENGTDVYGVKPKDYTGEDVTP 110

Qy 128 QNFYNVLLGNESGVTG-GNGKVKVSGPNDNFIYYADHAGAPGLIAMPTGDEVMAKDFNEV 186
Db 111 QNFVLVLRGDAEAVKGCGRKVLKSGPDHVFYFTDHGSGGILVFP-NEDLHVKDLIKT 169
Qy 187 LEKMHKRYKMYIYVEACSGSMFEGILKKNLNIYAVTAANSKSSWGVYCPESYPP 246
Db 170 THYIFKNMYKMYFYIEACSGSMNH-LPDNINVATTAAANPRESSYACYDE----- 223
Qy 247 PSEIGTCTFSISWLESDSLHDMKSETLEQQYHVVKRRVGSVDVPTSHVCRFGTEKML 306
Db 224 --KRSTYLGWDYSVNMEDSDVEDLTRETLLKQYHLVKSHT-----NTSHVMOYGNKTIS 276
Qy 307 KQYLSYIGRNPENDNFTTESFSS-----PISNSGLVNPDRIPILLYLQRIKQAPMGSL 361
Db 277 TMKVNQFQ-----MKRKASPVLPVPTHLDLTPSPDVLPTIMKRKLMT--NDL 325
Qy 362 ESK-----EAKKLLDEKHNKQIDOSITDILRLSVKQTNVNLNLTSTRTTGOPLVDDMDC 417
Db 326 ESRQLTEEIQRILDAH-----LIRGEVEQ-----LLSERA---PLTGH-SC 364
Qy 418 FKTLVNSFKNHC 429
Db 365 YPEVLLYFRTHC 376

RESULT 3

US-10-043-487-298
; Sequence 298, Application US/10043487
; Publication No. US20030055220A1
; GENERAL INFORMATION:
; APPLICANT: HYBRIGENICS
; APPLICANT: Pierre, LEGRAIN
; TITLE OF INVENTION: Protein-protein interactions between Shigella Flexneri polypep
; TITLE OF INVENTION: mammalian polypeptides
; FILE REFERENCE: B4778A
; CURRENT APPLICATION NUMBER: US/10/043,487
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/261,130
; NUMBER OF SEQ ID NOS: 561
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 298
; LENGTH: 319
; TYPE: PRT
; ORGANISM: Shigella Flexneri
US-10-043-487-298

Query Match 19.8%; Score 487; DB 9; Length 319;
Best Local Similarity 33.5%; Pred. No. 1e-33;
Matches 115; Conservative 64; Mismatches 124; Indels 40; Gaps 12;

Qy 133 VLLGNESGVTG-GNGKVKVSGPNDNFIYYADHAGAPGLIAMPTGDEVMAKDFNEVLEKMH 191
Db 2 VLRGDAEAVKGCGRKVLKSGPDHVFYFTDHGSGGILVFP-NEDLHVKDLNETIHYM 60
Qy 192 KKKYKMYIYVEACSGSMFEGILKKNLNIYAVTAANSKSSWGVYCPESYPPSEIG 251
Db 61 KKKYKMYIYVEACSGSMNH-LPDNINVATTAAANPRESSYACYDE-----KRS 112
Qy 252 TCGDPTFSISWLESDSLHDMKSETLEQQYHVVKRRVGSVDVPTSHVCRFGTEKMLKDYLS 311
Db 113 TYLGDWYSVNMEDSDVEDLTRETLLKQYHLVKSHT-----NTSHVMOYGNKTISTMKVM 167
Qy 312 SYIGRNPENDNFTTESFSS-----PISNSGLVNPDRIPILLYLQRIKQAPMGSL-ESKE 365
Db 168 QFQ-----MKRKASPVLPVPTHLDLTPSPDVLPTIMKRKLMT--NDLESRQ 216
Qy 366 AQKLLDEKHNKQIDOSITDILRLSVKQTNVNLNLTSTRTTGOPLVDDMDCFKTLVNSF 425
Db 217 LTEEIQRILDAHILKSVKRVKIVSLAASAEVQLLSERA---PLTGH-SCYPEALLHF 272
Qy 426 KNHC-----GATVHYGLKYTGALANICNMGVDVKQTVSAIEQAC 464

Db 273 RTHCFNMHSPTYEALRHLYVLNLCCKXPPLHRIKLSMDHVC 315

RESULT 4

US-10-114-170-257

; Sequence 257, Application US/10114170

; Publication No. US20030023075A1

; GENERAL INFORMATION:

; APPLICANT: Blattner, Frederick R.

; Burland, Valerie

; Perna, Nicole T.

; Plunkett, Guy

; Welch, Rod

; TITLE OF INVENTION: No. US20030023075A1el Sequences of E. coli O157

; NUMBER OF SEQUENCES: 265

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Quarles & Brady

; STREET: 1 South Pinckney Street

; CITY: Madison

; STATE: WI

; COUNTRY: US

; ZIP: 53701-2113

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 3.50 inch, 1.44Mb storage

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: word perfect 8.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/10/114,170

; FILING DATE: 01-Apr-2002

; CLASSIFICATION: <Unknown>

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 09/453,702

; FILING DATE: 03-DEC-1999

; APPLICATION NUMBER: 60/110,955

; FILING DATE: 04-DEC-1998

; ATTORNEY/AGENT INFORMATION:

; NAME: Seay, Nicholas J.

; REGISTRATION NUMBER: 27386

; REFERENCE/DOCKET NUMBER: 960296.95017

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (608) 251-5000

; TELEFAX: (608) 251-9166

; INFORMATION FOR SEQ ID NO: 257:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 3169 amino acids

; TYPE: amino acid

; STRANDEDNESS: <Unknown>

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; SEQUENCE DESCRIPTION: SEQ ID NO: 257:

US-10-114-170-257

Query Match 4.7%; Score 115.5; DB 9; Length 3169;

Best Local Similarity 21.2%; Pred. No. 2.3;

Matches 115; Conservative 83; Mismatches 188; Indels 157; Gaps 30;

QY 18 LIFSASRKTOLLNDN---DVESSDKSAK---GTRWAVLVAGSNE--YNYRHOADICH 68

Db 2496 IYGNHGAKEYTFLNSGSKTDIHLDRKNSDYDTDFRNIIFEYHTNIETIFSDNQGF-- 2553

QY 69 AQILRGKGLDENIIVFMDIAFSSFNPRPGVIINKPGE-----DVKGVKPDYTK 123

Db 2554 VISLNNATSAANINFRKNMTSLDSS---GSLIVLPAGDIYHSDIYK--MSRGRKSF 2608

QY 124 AVNQV-----NFNVLGNESGVTGGKVKVSGPN--DNIFTYYADHAGPLIAPPTG 175

Db 2609 KLVNEKRPDIDIINVAILETSYQ-----IKKIPNDDSDYILCLDN--PNLSS----- 2656

QY 176 DEVNAKDFNEV-----LEKMHK-----RKKYKNKMYIYVEACSGSMFE 213

Db 2657 ---YTLNFNDLSGYISSLWMDNIRGTSFTPFHKNTVNIAPNEKKYISLIGLDKLSFNIDVFR 2713

QY 214 GILK-KNLNIYAVTAANSKESWGVI-----CPE-----SYPPPPSEICTCLGD 256

Db 2714 QALEVKNKNSYKIS-----KFTWETYGDIVVSPEDRISHLELDGDFNYFSQP-ELDTPISD 2767

QY 257 TFSISW-----LESDSLH-----DMSKETLEQOYH--VVKRRVGSVDVPETSHVCRGTCKML 306

Db 2768 SFSYLYDNFQIVDSVDVHIKLLHLNRETQKITPHRIILKRYFIDSFAKTSITDR---EKNI 2824

QY 307 -----KDYLSYVIGRN-----PENDNFTFTESFSPISNSGLVNPDRDIPLL 347

Db 2825 YPVICDSPDHFTSDIYRHPFRIVLGNKTLIPSELVKFEISTSKYELSNMVDINNVIIVP-- 2882

QY 348 YLORKIQKAPMGSLKESKAEQKLLDEKNHRKQIDOSITDILRLSVKQTNV-----LN--- 399

Db 2883 -----QKTTKKNKLSIVSLNSNIKNDIVLSGVTGTGSKIFHLNNSG 2923

QY 400 --LLTSTRTTGCOPLVDDWDCFKTLVNSF--KNHCGATVHYGLKTYTGALANICNMGVDVKQ 455

Db 2924 DLLLTSTKTHGGVV---VIFKDFINNMWKNYLTITVPIDNKLSNDRINITPMTGKIQE 2980

QY 456 TVS 458

Db 2981 TVS 2983

RESULT 5

US-10-202-211-2

; Sequence 2, Application US/10202211

; Publication No. US20030003551A1

; GENERAL INFORMATION:

; APPLICANT: Itch, No. US20030003551A1uya

; TITLE OF INVENTION: METHOD FOR PRODUCING DIHYDROXYACETONE-3-PHOSPHATE

; FILE REFERENCE: 06501/038001

; CURRENT APPLICATION NUMBER: US/10/202,211

; CURRENT FILING DATE: 2002-07-22

; PRIOR FILING DATE: 1999-12-02

; NUMBER OF SEQ ID NOS: 4

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 2

; LENGTH: 580

; TYPE: PRT

; ORGANISM: Schizosaccharomyces pombe

US-10-202-211-2

Query Match 4.7%; Score 114.5; DB 9; Length 580;

Best Local Similarity 19.3%; Pred. No. 0.23; Mismatches 150; Indels 133; Gaps 20;

Matches 85; Conservative 72;

QY 19 IFSASRKTOLLNDNDESSDKSAKTRWAVLVAGSNEYNYRHOADICH---AVOILRK 75

Db 77 IFASPSSKQIYTGIKQVESE---AGT---LVICKN-----YTGDLILHFGMALEKORT 122

QY 76 GGLKDENIIVFMDIAF-----SSENPRPG-----VIINKPGEDVYKGVKPDYKAEVYN- 126

Db 123 AGKAAE--LIAVADDVSGVRKSKGVGRGLSGTVLVHKAAGAAAARGLP---LEAVTT 176

QY 127 -----VQNFYVNLGNESGVTGGKGVKVKSG--PNDNFIYYADHAGPG---LTMPTGD 176

Db 177 IAKAIDNLVSIAGSLAHVHPVHGHEPIAKEDMKHDELMGMIHNEPGCKRISPISID 236

QY 177 EVMAKDNEVLEKMHKRRKKNYKMYIYVEACSGSMFEIILKKNLNINIVAVTAANSKESWG 236

Db 237 DLIAQMLKQMLDOSDKDRAYVKI-----EG----- 261

QY 237 YCPESYPPPPSEICTCLGDTFTSISWLESDSLHDMKSKETLEQQYHVYKRRVGSVDVPETSH 296

Db 262 -----DDEVVLLMNLGGLSMLFEFSAISHKVKAEALAKYKINPVIRFAG--PFTTS 310

QY 297 VCRFG-----TEKMLKDYLSYIGRNPNENDNFTFTESFSSPISNSG--LVNPRDIP 346

Db 311 LNLGFGITLRTTDRV-----KVEGEYSLVLDLIDQPVVEAIGWPLCOPSDLK- 358

QY 347 LYLQRKIAPMGSLESKEAOKKL-----LDEKNHRKQIDQSITDILRLSVKQTNVLNLL 401
Db 359 -----SKNIGNSVIEGQDKVSPVTVDREKVRQAIIVNSMENLIIKAEPIITKF----- 407
QY 402 TSTRTTGQPLVDDWDCFKTL 421
Db 408 -----DTMAGDCGCGTTL 420

RESULT 6
US-09-738-626-3847
: Sequence 3847, Application US/09738626
: Publication No. US20020197605A1
: GENERAL INFORMATION:
: APPLICANT: NAKAGAWA, SATOSHI
: APPLICANT: NIZOGUCHI, HIROSHI
: APPLICANT: ANDO, SEIKO
: APPLICANT: HAYASHI, MIKIRO
: APPLICANT: OCHIALI, KEIKO
: APPLICANT: YOKOI, HARUHIKO
: APPLICANT: TATEISHI, NAOKO
: APPLICANT: SENOH, AKIHIRO
: APPLICANT: IKEDA, MASATO
: APPLICANT: OZAKI, AKIO
: TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
: FILE REFERENCE: 249-125
: CURRENT APPLICATION NUMBER: US/09/738,626
: CURRENT FILING DATE: 2000-12-18
: PRIOR APPLICATION NUMBER: JP 99/377484
: PRIOR FILING DATE: 1999-12-16
: PRIOR APPLICATION NUMBER: JP 00/159162
: PRIOR FILING DATE: 2000-04-07
: PRIOR APPLICATION NUMBER: JP 00/280988
: PRIOR FILING DATE: 2000-08-03
: NUMBER OF SEQ ID NOS: 7059
: SOFTWARE: Patent in ver. 3.0
: SEQ ID NO 3847
: LENGTH: 996
: TYPE: PRT
: ORGANISM: Corynebacterium glutamicum
US-09-738-626-3847

Query Match 4.4%; Score 107; DB 9; Length 996;
Best Local Similarity 19.9%; Pred. No. 2.3; Indels 104; Gaps 19;
Matches 90; Conservative 66; Mismatches 193;

QY 23 ESRKTOLLNDVSS-----DKSAG-----TRWAVLVAGSNEYNYRHOADICHAYQIL 73
Db 565 EAFNTSLV-DYDTTSWEDELNIAAGREGTEWL-----NGFYGDAAEQSWAESVA 617
QY 74 RKGGLK---DENI-----IVFYDDIAFSSENPRPG-----VIINKPDGEDVY 113
Db 618 ROGLKALVDANLEHIDARSVNSLKLFDAAEGRVNRVGRYGPYIERIVGTAEGEPEF 677
QY 114 K--GVPKDYTEAVNQ-----NFYNYLLGNESGVTGGNGKVY-----KSGPNDNFIYYA 162
Db 678 QRANLPEETTPDELTEVAEKLFPATQGGRELINPANGRMVVAKEGRFGP--YVIEQVT 735
QY 163 DHGAPGLIAMPTGDEVMA-----KDFNEVLEKMKHKKRYNKMVIYVEA 205
Db 736 DSERAG--AEAQAEVVAERKAEDBOATDGMRPKNWETKTAANOKEKRIINVEENLK 793
QY 206 CEGSMFEGTLKKNLNIYAVTAANSKSSWGVCYPPSEIGTCGLDFTFISLWLD 265
Db 794 PATASLFGMEPATVILEEALKLSLPREVGV-----DPSDNEVITAGNGRYGPLYKKG 847
QY 266 SLDLHDSKETLEOYHVVRVRSVDVPETSHVCREGTEKMLKDYLSYIGRNPENDNETF 325
Db 848 SDSRLNSE--EQIFVTLDARRIYAEPRCRAAAQPLQLGDNVSGKP-----MTV 901
QY 326 TESSFSPISNGLVNP-----DIP-----LLYLQRKIAPMGSLESKEA----- 367

Db 902 KGRFGYVTDGTTNASLRKGDVPESLTDARANELLSERRAKEADGAPAKKTTSTKTKTA 961
QY 368 -----KKLDEKNHRKQIDQSITDILRLSVKQ 394
Db 962 AKTTAKKTTAKTKVRKAPPKTKTNVVKAGAKK 994

RESULT 7
US-09-815-242-5140
: Sequence 5140, Application US/09815242
: Patent No. US20020061569A1
: GENERAL INFORMATION:
: APPLICANT: Haselbeck, Robert
: APPLICANT: Ohlsen, Karl L.
: APPLICANT: Zyskind, Judith W.
: APPLICANT: Wall, Daniel
: APPLICANT: Trawick, John D.
: APPLICANT: Carr, Grant J.
: APPLICANT: Yamamoto, Robert T.
: TITLE OF INVENTION: Identification of Essential Genes in
: FILE REFERENCE: ELITRA.011A
: CURRENT APPLICATION NUMBER: US/09/815,242
: CURRENT FILING DATE: 2001-03-21
: PRIOR APPLICATION NUMBER: 60/191,078
: PRIOR FILING DATE: 2000-03-21
: PRIOR APPLICATION NUMBER: 60/206,848
: PRIOR FILING DATE: 2000-05-23
: PRIOR APPLICATION NUMBER: 60/207,727
: PRIOR FILING DATE: 2000-05-26
: PRIOR APPLICATION NUMBER: 60/242,578
: PRIOR FILING DATE: 2000-10-23
: PRIOR APPLICATION NUMBER: 60/253,625
: PRIOR FILING DATE: 2000-11-27
: PRIOR APPLICATION NUMBER: 60/257,931
: PRIOR FILING DATE: 2000-12-22
: PRIOR APPLICATION NUMBER: 60/269,308
: PRIOR FILING DATE: 2001-02-16
: NUMBER OF SEQ ID NOS: 14110
: SOFTWARE: FastSEQ for Windows Version 4.0
: SEQ ID NO 5140
: LENGTH: 568
: TYPE: PRT
: ORGANISM: Pseudomonas aeruginosa
US-09-815-242-5140

Query Match 4.2%; Score 103; DB 10; Length 568;
Best Local Similarity 21.1%; Pred. No. 2.2; Indels 108; Gaps 16;
Matches 65; Conservative 36; Mismatches 99;

QY 39 DKSAGTRWAVLVAGSNEYNYRHOADICHAYQILRKGLKLDENIIVFYDDIAFSSENP 98
Db 62 DODAK-----LLVPFDNDY-----DNIVASAYQGTLMAREE 93
QY 99 RGVVIINKPDGEDVYKGVPKDYTKAVNV--QNFYNYLLGNESGVTGGNGKVYKSG----- 152
Db 94 RWNLIQAK-DGK-----VLRDDIGEALSLTLPNLYGFVRDGYGVVDGQCKEYQAPRFD 147
QY 153 --PND-NIFYYADHGAPGLIAMPTGDEVMAKDFNEVLEKMKHKKRYNKMVIYVEACESG 209
Db 148 IYPSANEFIIYETDGRGTL-----DAKGKQLTEAL-----YDTTLVNGSVAEHG 193
QY 210 SMFEGILKKNLNIYAVTAANSKSSWGVCYPPSEIGTCGLDFTFISLWLDSDLIH 269
Db 194 GLI-----SAERGEKW-----IINLATGEOKAVAVESLGLIH 226
QY 270 D--MSKETLEOYHVVKRR---VGSVDVPETSHVCREGTEKMLKDYLSYIGRNPENDNET 324
Db 227 DGVMSASVIGKGSOLVDAKGDVVODG-----KSY--DYLG--TPANGLVA 267
QY 325 FTESFSSP 332

Db 268 PREKYDSP 275

RESULT 8

US-09-984-130-43
 ; Sequence 43, Application US/09984130
 ; Publication No. US20030055231A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Ni et al.
 ; TITLE OF INVENTION: 12 Human Secreted Proteins
 ; FILE REFERENCE: PF489P2
 ; CURRENT APPLICATION NUMBER: US/09/984,130
 ; CURRENT FILING DATE: 2001-10-29
 ; PRIOR APPLICATION NUMBER: 60/243,792
 ; PRIOR FILING DATE: 2000-10-30
 ; PRIOR APPLICATION NUMBER: 09/836,353
 ; PRIOR FILING DATE: 2001-04-18
 ; PRIOR APPLICATION NUMBER: 60/198,407
 ; PRIOR FILING DATE: 2000-04-19
 ; PRIOR APPLICATION NUMBER: PCT/US99/25031
 ; PRIOR FILING DATE: 1999-10-27
 ; PRIOR APPLICATION NUMBER: 60/105,971
 ; PRIOR FILING DATE: 1998-10-28
 ; NUMBER OF SEQ ID NOS: 149
 ; SOFTWARE: Patent In Ver. 2.0
 ; SEQ ID NO 43
 ; LENGTH: 1034
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-984-130-43

Query Match 4.2%; Score 102; DB 9; Length 1034;

Best Local Similarity 20.8%; Pred. No. 6.4;

Matches 92; Conservative 56; Mismatches 156; Indels 138; Gaps 22;

```

Qy 71 QILRGGKLDENIIVFMYDDIAFSSSENPGRPGVIINKPGEDVYK---GVPRDYTKVAVNV 127
   :||| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 252 EAFQGGGRGKAKKVMIVITD-GESHDSPDLEKVIQOQSERDNTRYAVAVLGYNRRGINP 310
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Qy 128 QNFYVNLGNESGVTGGNGKVPKNDNFIYYADHGAPGLIAMPTGDEVMAKDFNEVL 187
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 311 ETFLNEI-----KYIASDPDDKHFFENVTDAAALDKDIVDALGDRIFS-----L 352
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Qy 188 EKMHRKKYKMKVIYVEACESG-----SMFEGILKNLIYAVTAANSKESGVVYCP--E 241
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 353 EGTNK-----NETSFGLEMSQTFSSHVVEDGVLLGAVGAYDWNGAVLKETSAGKVIPLRE 408
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Qy 242 SY-----PPPPSEIGICLGDFTSISWLESDSLHDMSKETLEQOYHVVKRRVG-----SDVPE 293
   || :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 409 SYLKEFPEELKNHGAYLGTYVT-----SVVSRQGRVYVAGAPR 447
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Qy 294 TSH-----VCRFGT-----EKMLKDYLSYIGRNPNENDNFTTFESFSPISNSGLVNP 341
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 448 FNHTGKVILFTMHNRRSLTIHQAMRGOQIGSYFG-----SEITSVDIDGDGVT-- 495
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Qy 342 RDIPLLYLQRIKQAPMGSLKESKEAQK-----KLLDEKNHRK-QIDQSIT 385
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 496 -DVLIV-----GAPMYFNEGREGKVVYVELRQNRFFVYNGTLKDSHSYQNAFGSSIA 547
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Qy 386 DILRLSVKQTNVNLTLSTRTTGPPLVDWDCFKTLVNSFKNHCCAT-VHYGLKVTGALA 444
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 548 SVRLNQDSYNDV-----VVGAPLED-----NHAGAIYIFHG--FRGSIL 585
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Qy 445 NICNMGVDVKQTVSAIEQAQCSM 466
   || :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 586 K-----TPKQRITASELATGL 601
   || :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|

```

RESULT 9

US-09-984-130-35
 ; Sequence 35, Application US/09984130
 ; Publication No. US20030055231A1
 ; GENERAL INFORMATION:

; APPLICANT: Ni et al.
 ; TITLE OF INVENTION: 12 Human Secreted Proteins
 ; FILE REFERENCE: PF489P2
 ; CURRENT APPLICATION NUMBER: US/09/984,130
 ; CURRENT FILING DATE: 2001-10-29
 ; PRIOR APPLICATION NUMBER: 60/243,792
 ; PRIOR FILING DATE: 2000-10-30
 ; PRIOR APPLICATION NUMBER: 09/836,353
 ; PRIOR FILING DATE: 2001-04-18
 ; PRIOR APPLICATION NUMBER: 60/198,407
 ; PRIOR FILING DATE: 2000-04-19
 ; PRIOR APPLICATION NUMBER: PCT/US99/25031
 ; PRIOR FILING DATE: 1999-10-27
 ; PRIOR APPLICATION NUMBER: 60/105,971
 ; PRIOR FILING DATE: 1998-10-28
 ; NUMBER OF SEQ ID NOS: 149
 ; SOFTWARE: Patent In Ver. 2.0
 ; SEQ ID NO 35
 ; LENGTH: 1189
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-984-130-35

Query Match 4.2%; Score 102; DB 9; Length 1189;

Best Local Similarity 20.8%; Pred. No. 7.9;

Matches 92; Conservative 56; Mismatches 156; Indels 138; Gaps 22;

```

Qy 71 QILRGGKLDENIIVFMYDDIAFSSSENPGRPGVIINKPGEDVYK---GVPRDYTKVAVNV 127
   :||| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 252 EAFQGGGRGKAKKVMIVITD-GESHDSPDLEKVIQOQSERDNTRYAVAVLGYNRRGINP 310
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Qy 128 QNFYVNLGNESGVTGGNGKVPKNDNFIYYADHGAPGLIAMPTGDEVMAKDFNEVL 187
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 311 ETFLNEI-----KYIASDPDDKHFFENVTDAAALDKDIVDALGDRIFS-----L 352
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Qy 188 EKMHRKKYKMKVIYVEACESG-----SMFEGILKNLIYAVTAANSKESGVVYCP--E 241
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 353 EGTNK-----NETSFGLEMSQTFSSHVVEDGVLLGAVGAYDWNGAVLKETSAGKVIPLRE 408
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Qy 242 SY-----PPPPSEIGICLGDFTSISWLESDSLHDMSKETLEQOYHVVKRRVG-----SDVPE 293
   || :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 409 SYLKEFPEELKNHGAYLGTYVT-----SVVSRQGRVYVAGAPR 447
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Qy 294 TSH-----VCRFGT-----EKMLKDYLSYIGRNPNENDNFTTFESFSPISNSGLVNP 341
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 448 FNHTGKVILFTMHNRRSLTIHQAMRGOQIGSYFG-----SEITSVDIDGDGVT-- 495
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Qy 342 RDIPLLYLQRIKQAPMGSLKESKEAQK-----KLLDEKNHRK-QIDQSIT 385
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 496 -DVLIV-----GAPMYFNEGREGKVVYVELRQNRFFVYNGTLKDSHSYQNAFGSSIA 547
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Qy 386 DILRLSVKQTNVNLTLSTRTTGPPLVDWDCFKTLVNSFKNHCCAT-VHYGLKVTGALA 444
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 548 SVRLNQDSYNDV-----VVGAPLED-----NHAGAIYIFHG--FRGSIL 585
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Qy 445 NICNMGVDVKQTVSAIEQAQCSM 466
   || :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 586 K-----TPKQRITASELATGL 601
   || :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|

```

RESULT 10

US-09-991-258-3
 ; Sequence 3, Application US/09991258
 ; Patent No. US20020141975A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Olmsted, Robert
 ; APPLICANT: Keith, Paula
 ; APPLICANT: Dryga, Sergey
 ; APPLICANT: Caley, Ian
 ; APPLICANT: Maughan, Maureen
 ; APPLICANT: Johnston, Robert
 ; APPLICANT: Davis, Nancy
 ; APPLICANT: Swanstrom, Ronald

```

; TITLE OF INVENTION: ALPHAVIRUS VECTORS AND VIROSOSES WITH MODIFIED HIV GENES FOR USE
; FILE REFERENCE: 0113 000103
; CURRENT APPLICATION NUMBER: US/09/991,258
; PRIOR FILING DATE: 2001-11-16
; PRIOR APPLICATION NUMBER: 09/902,537
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: 60/216,995
; PRIOR FILING DATE: 2000-07-07
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 2492
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: No. US20020141975A1e =
; OTHER INFORMATION: synthetic construct
US-09-991-258-3

Query Match      4.1%: Score 100.5; DB 10; Length 2492;
Best Local Similarity 19.7%: Pred. No. 31;
Matches 77; Conservative 59; Mismatches 136; Indels 119; Gaps 18;

QY 17 LLIFSAESRKTQLNDNDVSSDSKAKGTRWAVLVAGSNYYNR-HOADIHAYQILRK 75
DB 1293 LFVFTGYDKRARTHPYKLSST-----LTNIYTGSRLEHACGAPSYHVR- 1337

QY 76 GGLKDNIIIVMYDDIAPFSSNPRPGVIIN-----KPDG---EDVYKGVKDYTKEAVN 126
DB 1338 -----GDIATATE-----GVIINAANSKGGPGGVCALYKFPESFDLOPIE 1380

QY 127 VQNFYNVLGNESGTGNGKVVKS-GPNDNIFYIYADHAGAPGLIAMPDGTGVMAKDFNE 185
DB 1381 V-----GRARLVKGAAKHIIHAGVFNFN-----KVSEVEDGDKLAEAY-E 1419

QY 186 VLEKMKRKKYKMKVIYVEACESGMPEG---ILKKNLNIYAVTAANSKSSWGVCPE- 241
DB 1420 STAKIVNDNNYKSAI-----PLLSIGISGNKDRLTQSLN-HLLFALDITDADVAIYCRDK 1475

QY 242 -----SYPPPPSEI-----GTCLGDTFFSISWL 263
DB 1476 KWMTLKAVARREAVEICISDDSSVTEPDALVVRHPKSSLAGRGKYSTGDKTFSYL 1535

QY 264 EDSLDHMSKETLEQOYHVVRVGVSDVPETSHVCRFGTEKMKLDYLLSSYIGRNPNDFN 323
DB 1536 EGTKEHQAKDAEIN---AMWPVATEANE---QVCMT-----ILGESMSSTRSKCPVEESE 1586

QY 324 TTFESFSPISNSGLVNPDPRIPLLYLQRIQ 354
DB 1587 ASTPPSTLPLCLCIHAMTPERVORLKASRPEQ 1617

RESULT 11
US-09-816-685-4
; Sequence 4, Application US/09816685
; Patent No. US20020053091A1
; GENERAL INFORMATION:
; APPLICANT: CHANDRAMOULISWARAN, Ishwar et al
; TITLE OF INVENTION: ISOLATED HUMAN G-PROTEIN COUPLED
; TITLE OF INVENTION: RECEPTORS, NUCLEIC ACID MOLECULES ENCODING HUMAN GPCR
; TITLE OF INVENTION: PROTEINS, AND USES THEREOF
; FILE REFERENCE: CL000856
; CURRENT APPLICATION NUMBER: US/09/816,685
; CURRENT FILING DATE: 2001-03-26
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 828
; TYPE: PRT
; ORGANISM: Carassius auratus
US-09-816-685-4

```

```

Query Match      4.1%: Score 100; DB 10; Length 828;
Best Local Similarity 20.5%: Pred. No. 6.9; Indels 92; Caps 17;
Matches 72; Conservative 50; Mismatches 138;

QY 15 HALLIFSAESRKTQLNDNDVSSDSKAKGTRWAVLVAGSNYYNRHQADI 66
DB 168 HAMV-----QULKDN-----KWT-WVGIITLDGDIYRSAMESPVKHTEREKI 208

QY 67 CHAYQILRKGGLKDE-NIIVFMYDDIAPFSSNPRPGVIINPKPDGED---VYKG-----VP 117
DB 209 CVAFKVIPLDPSLADBEQKLNHINETVDIIIEKNTKVNVVVSFAKSSOMKLLYEGULSRNVP 268

QY 118 KDYTKEA-----VNVQNFYNVLGNESGVTGNGKVVKSGPNDNIFYIYADH 164
DB 269 KKVWVASDNWSTSKNLIKDVNLSDIGNIL-----GFTFKSG-NVTAFLQY--- 313

QY 165 GAPGLIAMPDGTGVMAKDFNEVLEKMKRKKYKMKVIYVEACESGMPEGILKKNLNIYA 224
DB 314 ----LKDLKFGSE--AKMNSFLEEFKLPEIGNAANAQVQIKNTHLDVMSVQMAVSA 367

QY 225 VTAANSKSSWGVCPEYPPPPSEIGTCLGDTFFSISWLESDSLDHMSKETLEQOYHVVK 284
DB 368 IAKAVVE-----LCVERQCKTPSAIQ-----PW---ELLQLRNVTFEKEGVMYN 409

QY 285 RRVGSDVPETSHVCRFGTEKMKLDYLLSSYIGRNPNDFNFTTFESFSPISN 335
DB 410 PDANGDINLGYDVLWDDDESEKNDIIAEYV---PSNSSFTFTRKNLSNIEN 458

RESULT 12
US-10-125-540-313
; Sequence 313, Application US/10125540
; Publication No. US20030059875A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT214CI
; CURRENT APPLICATION NUMBER: US/10/125,540
; CURRENT FILING DATE: 2002-04-19
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 646
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 313
; LENGTH: 707
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-125-540-313

Query Match      4.0%: Score 99; DB 9; Length 707;
Best Local Similarity 20.6%: Pred. No. 6.7;
Matches 91; Conservative 57; Mismatches 156; Indels 138; Caps 22;

QY 71 QILRKGGLKDNIIIVFMYDDIAPFSSNPRPGVIINPKPDGEDVYK---GVPKDYTKEAVN 127
DB 268 EAFQKGRKGRKAKKVMIVITD-GESHDSPDLEKVIQQSERDNNVTRYAVAVLGYNNRGINP 326

QY 128 ONFYNVLGNESGVTGNGKVVKSGPNDNIFYIYADHAGAPGLIAMPDGTGVMAKDFNEVL 187
DB 327 ETFLENEI-----KYIASDPDDKHFFNVTDEAALKDIVDALGDRIFS-----L 368

QY 188 EKMHRKKYKMKVIYVEACESG---SMFEGILKKNLNIYAVTAANSKSSWGVCPE--E 241
DB 369 EGTNK-----NETSFLEMSQTGFSSHVVEDGVLLGAVGAYDNGAVLKETSAGKVIPLRE 424

QY 242 SY-----PPPPSEIGTCLGDTFFSISWLESDSLDHMSKETLEQOYHVVRVVG-----SDVPE 293
DB 425 SYLKEFFPEELKNHGYALGYTVT-----SVMSRGRGVVYVAGAPR 463

QY 294 TSH---VCRFGT-----EKMCLKYLLSSYIGRNPNDFNFTTFESFSPISNSGLVNP 341
DB 464 FNHTCKVILFTMHNHNRSLTIHOAMRGQOIGSYFG-----SEITSDVIDGDGVT-- 511

QY 342 RDIPLLYLQRIQKAPMGSLSEKSAQK-----KLLDEKNHRK-QIDOSIT 385

```

```
Db 512 -DVLV-----GAPMFNEGRGRGVYVELRQNLVFNGLTKDSHSONARFGSSIA 563
Qy 386 DILRLSVKQTNVNLSTRTTGOPLVDDWDCFKTLNFKSNHCGAT-VHYGLKYTGALA 444
Db 564 SVRLNODSYNDV-----VVGAPLED-----NHAGAIYIFHG--FRGSIL 601
Qy 445 NICMGMVDVQTVSAIQACSM 466
Db 602 K-----TPKQIRITASELATGL 617

RESULT 13
US-09-764-870-313
; Sequence 313, Application US/09764870
; Patent No. US20020042386A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PTZ14
; CURRENT APPLICATION NUMBER: US/09/764,870
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 646
; SOFTWARE: PatentIn ver. 2.0
; SEQ ID NO 313
; LENGTH: 707
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-870-313

Query Match 4.0%; Score 99; DB 10; Length 707;
Best Local Similarity 20.6%; Pred. No. 6.7;
Matches 91; Conservative 57; Mismatches 156; Indels 138; Gaps 22;

Qy 71 QILRKGGLKDNIIYFMVDDIAFSSNPRGVIIINKPDGEDVYK---GVPKDYTKVAVV 127
Db 268 EAFQGGGRGAKKVMVITD--GESHSDPLEKVITQQSERDNVTYAVVLGYNNRRGINP 326
Qy 128 QNFYNVLIGNESGVTGGNGKVKSGPNDNIFYADHAGPLIAMPTGDEVMKDFNEVL 187
Db 327 ETFLNEI-----KYIASDPDKHFFNVYDDEALKDVIDALGDRIFS-----L 368
Qy 188 EKMHRKKYKMWIYVACESG-----SMFEGILKKNLIYAVTAANSKESGWYCP--E 241
Db 369 EGTNK-----NETSFGLEMSQTGFSSHVEDGVLLGAVGDWNGAVLKETSGAKVIPLE 424
Qy 242 SY-----PPPSEIGTCIGDTFISWLESDSLHDSKETLEQQYHVVKRYG----SDVPE 293
Db 425 SYLKEFPEELKNHGAYLGYVT-----SVMSRQGRVYVAGAPR 463
Qy 294 TSH-----YCREGT-----EKMVKDYLSSYIGRNPENDNFTFTSFSPISNSGLVNP 341
Db 464 FNHTGKVLFTMHNRSITIHQARMGOOIGSYFG-----SEITSVDIDGDVY-- 511
Qy 342 RDIPLLYLQRIQKAPMGSKESKAQK-----KLLDEKNIRK-OIDQSIT 385
Db 512 -DVLV-----GAPMFNEGRGRGVYVELRQNLVFNGLTKDSHSONARFGSSIA 563
Qy 386 DILRLSVKQTNVNLSTRTTGOPLVDDWDCFKTLNFKSNHCGAT-VHYGLKYTGALA 444
Db 564 SVRLNODSYNDV-----VVGAPLED-----NHAGAIYIFHG--FRGSIL 601
Qy 445 NICMGMVDVQTVSAIQACSM 466
Db 602 K-----TPKQIRITASELATGL 617
```

```
RESULT 14
US-08-945-749-3
; Sequence 3, Application US/08945749
; Patent No. US2002013880A1
; GENERAL INFORMATION:
```

```
; APPLICANT: GARDNER, Richard C
; APPLICANT: MACDIARMID, Colin W
; APPLICANT: HAY, Robert J
; APPLICANT: Auckland Uniservices Limited
; APPLICANT: New Zealand Pastoral Agriculture Research Institut
; TITLE OF INVENTION: ALUMINUM RESISTANCE GENE
; FILE REFERENCE: 08/945,749
; CURRENT APPLICATION NUMBER: US/08/945,749
; CURRENT FILING DATE: 1998-01-12
; EARLIER APPLICATION NUMBER: PCT/NZ96/0035
; EARLIER FILING DATE: 1996-05-01
; EARLIER APPLICATION NUMBER: NZ 272039
; EARLIER FILING DATE: 1995-05-01
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 969
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-08-945-749-3

Query Match 4.0%; Score 98.5; DB 8; Length 969;
Best Local Similarity 18.6%; Pred. No. 12;
Matches 87; Conservative 70; Mismatches 169; Indels 143; Gaps 19;

Qy 30 LINDNDV-----ESSDKSAKGRMAVLVAGSN---EYNYRHOA-----DIC 67
Db 207 VNDNSKRRKKRGSDSSNKNKSTS-----SDSNDDEDEYNSRPSLSNNSLDDVC 261
Qy 68 -----HAYQILRKGLKDNIIYFMVDDIAFSSNPRGV 102
Db 262 LVLDDEGEVPAWPCDCTVLEFSKEETERLSQAIODAEAFHFOYDEDEEDGTSNEDGI 321
Qy 103 IINKPCDGEDVYKVPKDYTKVAVNVQNFYN-----VLLGNESGVVTGGN 145
Db 322 LFSKPIVTNI--DYPELGNRRVNETENLKNRPLRPKRIAPHLLIQRPMVLCNS--TKDS 377
Qy 146 GKVVKSGPNDNIF---IYYADH-----GAPGLIAMPT----- 174
Db 378 KSRIQSGQLQDNLVGRNIQYPPHIIISNNEPFRFYERVDLDSVHSPTISGLLQPGKF 437
Qy 175 GDEVMKDFNEVLEKMKRKKYKMWIYVACESGSMFEGILKKNLIYAVTAANSKES 234
Db 438 ODUFVASIYSDNSAGHIKTHPNSTPTGIKA-ETVSQLOGLTAKN-----PSTLSS 487
Qy 235 WGVYCPESYPP-----PPSEIGTCIGDTFISWLEDSOL---HDMSKETLEQQYHVVK 284
Db 488 MSVANIEDVPPFWLDVSNPTEEBEMKILSKAFGIHPLTTEDIFLGEVREKVELFRDYILIC 547
Qy 285 RVGSDVPETSHVCRFCTEKMLKDYLSYIGRNPENDNFTFTSFSPISNSGLVNP-- 343
Db 548 FR-SEDIVAEKHVRRRKRKEQESATL-----DHESISRRKRSQAYGATMSNESNANNNS 600
Qy 344 -----IPLLQLRKIQKAPMGSKESKAQKLLDEKNHNRKOIDOS 383
Db 601 TSNASRSKWLPILRARRRSSANTTTNTSSSYKRRV--KSEKMKKEEN 647
```

```
RESULT 15
US-09-759-508B-2
; Sequence 2, Application US/09759508B
; Publication No. US20020182599A1
; GENERAL INFORMATION:
; APPLICANT: Fishman, Mark C.
; TITLE OF INVENTION: Methods for Diagnosing and Treating Heart Disease
; FILE REFERENCE: 00786/381002
; CURRENT APPLICATION NUMBER: US/09/759,508B
; CURRENT FILING DATE: 2001-01-12
; PRIOR APPLICATION NUMBER: US 60/175,787
; PRIOR FILING DATE: 2000-01-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
```

```

; LENGTH: 26926
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-759-508B-2

Query Match      4.0%; Score 98.5; DB 9; Length 26926;
Best Local Similarity 19.0%; Pred. No. 1.5e+03;
Matches 101; Conservative 81; Mismatches 152; Indels 197; Gaps 28;

Qy 20 FSAERSRKTOLLNDNDVSSDKSAGTRWAVLVAGSNEYNYRHQADICHAYQILRKGLK 79
Db 19621 YTVYKKS---DQDKWTSQSLRGTEYTGITGTAEVFRVKS-----VNVKGAS 19669
Qy 80 D-----ENIIVMYDD-----IAFSSENPRPGVIINKPG 109
Db 19670 DPSOSSDQIAKEREEPLDIOSEMRKTLIVKAGASFTMTVPFRG-RVPNVVLWSKPT 19728
Qy 110 -----EDVYKGVPKDYKQAVNVQNFYNYVLLGNESGVTGGNGKVYK 150
Db 19729 DLRTRAYVDTTDSRTSLTIENANRNDGKYTLTIQNVLSAASLT-----VVK 19776
Qy 151 ----SGPNDNIFIYYA-----DHGAPGLIAMPTGDEVMAKDFNEVLEKMHK 192
Db 19777 VLDTPGPTNITVDVTKESAVLSWDVPENDGGAP-----VKNYH--IEKREA 19822
Qy 193 RKK-----YKNMWIYVEACSGSMFEGILKKNLIYAVTAANS-----KESMGVY 238
Db 19823 SKKAWSVTNNCNRLSYKVTNLQEGAIY-----YFRVSGENEFVGIPTAEKGVK 19873
Qy 239 CPESYPPPPSEIG--TCLGDTFSISWLESDLDHMSKETLEQQYHVVKRVGSDVPETSH 296
Db 19874 ITEK-PSPPPEKLGVTISKDSVSLTLKPE--HDGGSRIV---HYVVE----- 19915
Qy 297 VCREGTEKMLKDYLSYIGRNP-----ENDNFTTFESFSPISNGLVNPRIPLLY 348
Db 19916 ----ALEKGQKNWVKCAVAKSTHHVVSGLRSENSEY-FFRVFAE--NOAGLSDPRELLPV 19968
Qy 349 LQRIQKAPMGSLESKEAOKKLLDEKNHRKQIDQISITDILRLSVKQTNVNLTLTSTRTG 408
Db 19969 LIKE-----QLEPPE-----IDMKNF-----PSHTVVVVRAG-----SNLKVDIPISG 20005
Qy 409 OPLVDDWDCEFTLVNSPKNHCAGATVHYGLKYGTCALANICNMGVGVKQTVSA 459
Db 20006 KPLP-----KVTLSRDGVPLKATMRNTEITAE-----NLINLRQSVTA 20045

```

Search completed: May 27, 2003, 15:36:36
Job time : 76 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 27, 2003, 15:24:54 ; Search time 45 Seconds
(without alignments)
995.525 Million cell updates/sec

US-09-934-066-2

Perfect score: 2454

Sequence: 1 MSSPLGHFQILVFLHALLIF.....CNMGVDVKQTVSAIEQAQCSM 466

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR73.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1268.5	51.7	493	2 S49175	legumain (EC 3.4.22.34)
2	1262.5	51.4	494	2 S51117	cysteine proteinase
3	1246	50.8	494	2 T05302	vacuolar processin
4	1243.5	50.7	484	2 T12043	probable legumain
5	1230.5	50.1	478	2 T02629	vacuolar processin
6	1148	46.8	484	2 S60050	vacuolar processin
7	1125	45.8	536	2 C96652	protein F23N19.7
8	1115.5	45.5	493	2 T12044	probable legumain
9	1109	45.2	497	2 J02387	vacuolar processin
10	1080	44.0	475	2 JX0344	legumain (EC 3.4.22.34)
11	1071	43.6	495	2 T07132	cysteine proteinase
12	1054	43.0	503	2 T10944	cysteine proteinase
13	790	32.2	429	2 A60145	hemoglobinase (EC 3.4.21.1)
14	782.5	31.9	462	2 T19231	probable cysteine
15	768	31.3	423	2 S31908	hemoglobinase - fl
16	326.5	13.3	326	2 T13411	hypothetical prote
17	321	13.1	411	2 S59796	probable membrane
18	292.5	11.9	380	2 T40853	probable cysteine
19	276	11.2	322	2 T24525	hypothetical prote
20	218	8.9	428	2 T00731	hypothetical prote
21	124	5.1	1234	2 T31623	hypothetical prote
22	118	4.8	1650	2 T18444	hypothetical prote
23	116	4.7	522	2 JA0072	hypothetical prote
24	115.5	4.7	3169	2 T00296	hypothetical prote
25	115	4.7	872	2 JC7380	toxin B - Escheric
26	114.5	4.7	580	2 T43310	DNA-directed DNA p
27	112.5	4.6	1440	1 SVHUOT	glycerone kinase
28	112	4.6	4981	2 T14899	multifunctional am
29	111.5	4.5	920	2 T41030	hypothetical prote
					conserved hypothet

30 111.5 4.5 1164 2 S46769
31 111.5 4.5 2401 2 T28676
32 111 4.5 706 2 T08313
33 109.5 4.5 761 2 B97163
34 109 4.4 507 2 A71622
35 109 4.4 786 2 T18469
36 108 4.4 635 2 F90551
37 108 4.4 679 2 G71615
38 107 4.4 774 2 T00488
39 106.5 4.3 452 2 F96981
40 106.5 4.3 1059 1 S61311
41 105.5 4.3 719 2 S44237
42 105.5 4.3 800 2 S53079
43 105 4.3 694 2 T10565
44 104.5 4.3 451 2 S44161
45 104.5 4.3 1252 2 H97178

ALIGNMENTS

RESULT 1

S49175

legumain (EC 3.4.22.34) precursor [similarity] - spring vetch

C:Species: Vicia sativa (spring vetch, tare)

C>Date: 16-Feb-1995 #sequence_revision 12-May-1995 #text_change 21-Jul-2000

C:Accession: S68984; S49175

R:Becker, C.; Shutov, A.D.; Nong, V.H.; Senyuk, V.I.; Jung, R.; Horstmann, C.; Fische

Eur. J. Biochem. 228, 456-462, 1995

A:Title: Purification, cDNA cloning and characterization of proteinase B, an asparagi

A:Reference number: S68984; MUID:95220376; PMID:7705362

A:Accession: S68984

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-493 <BE2>

A:Cross-references: EMBL:Z34899; NID:9510357; PIDN:CAA84383.1; PID:9510358

C:Superfamily: legumain

C:Keywords: cysteine proteinase; glycoprotein; hydrolase

F:1-24/Domain: signal sequence #status predicted <SIG>

F:25-51/Domain: propeptide #status predicted <PRO>

F:52-493/Product: legumain #status predicted <MAT>

F:147,295,331/Binding site: carboxylate (Asn) (covalent) #status predicted

Query Match 51.7% Score 1268.5; DB 2; Length 493;

Best Local Similarity 53.7%; Pred. No. 5.9e-80;

Matches 235; Conservative 70; Mismatches 122; Indels 11; Gaps 3;

Qy 31 NDNDVSSDKSAKGRVAVLVAGSNEYNYRHQADICHAYQILRKGGKLDENIIVFYDD 90

Db 48 NDDDFE-----GTRWAILLAGSNGYNYRHQSDVCHAYQLLRKGGSKKEENIIVFYDD 100

Qy 91 IAFSENPRPGVIINKPDGEDVYKGVKDYTKAEAVNVONVNLGNSGVTVGGNGKVK 150

Db 101 IASNEENPRPGVIINKPDGDDVYAGVPRDYTGAEVHADNFYAAALGNKSAITGGSGKVV 160

Qy 151 SGPNDNFIYYADHAGPLIAMPTGTDEVMKDFNEVLEKMKRKKYKMKVIYVEACES 210

Db 161 SGPNDHIEVYTDHGGPGVGLMPGVLYASDLNVLKHKHAGSYKSLVFLYECES 220

Qy 211 MFEGILKKNLNIYAVTAANSKSSWGVCPSYPPPPSEIGTCIGDTTSSIWLESDLD 270

Db 221 IFEGLLPDLNIAIYATASNAEESWGVCYCPGDKPPPEYSTCLGLDLYSIAMNDESEVN 280

Qy 271 MSKETLEQQYHVKKRRVSDVPETSHVCRFGTEKMLKDYLSYICRNPNDNDTFTES 330

Db 281 LQTESLQQYKLVKNTISE-PYGSHVMEYDGLGKNDLYOYLTGNFANDNNSFVDETE 339

Qy 331 SPI---SNSGLVNPRIPLLYLQRIQKAPMGSLKESQAQKLLDEKHNKQIDQSI 387

Db 340 NSLKRTSAAVNQDADLIHFWEKFRAPESGSSOKNEAKQVLEAMSHRKHIDNSVK 399

Qy 388 LRLSVKQTNVLNLTSTRTTGOPLVDDDCFKTLVNSKFNHCGATVHYGLKTYTGAL 447

Db 388 LRLSVKQTNVLNLTSTRTTGOPLVDDDCFKTLVNSKFNHCGATVHYGLKTYTGAL 447

R:Bevan, M.; Terry, N.; Ardiles, W.; Buysshaert, C.; Dasseville, R.; De Clerck, R.;
ewes, H.W.; Mayer, K.F.X.; Schueller, C.
submitted to the Protein Sequence Database, October 1998
A:Reference number: Z15407
A:Accession: T05302
A:Molecule type: DNA
A:Residues: 1-494 <BEV>
A:Cross-references: EMBL:AL031804
A:Experimental source: cultivar Columbia; BAC clone F36P21
C:Genetics:
A:Map position: 4
A:Introns: 76/3; 131/3; 184/1; 212/3; 279/2; 295/3; 365/3; 434/3
A:Note: F26P21.60
C:Superfamily: legumin
C:Keywords: cysteine proteinase; glycoprotein; hydrolase
F:1-27/Domain: signal sequence #status predicted <SIG>
F:28-53/Domain: propeptide #status predicted <PRO>
F:54-49/Product: vacuolar processing enzyme isozyme gamma #status predicted <MAT>
F:336/Binding site: carboxylate (Asn) (covalent) #status predicted

. Query Match 50.8% Score 1246; DB 2; Length 494;
Best Local Similarity 53.6%; Pred. No. 2,1e-76;
Matches 233; Conservative 121; Mismatches 121; Indels 10; Gaps 4;

Qy	36	ESSDKSAKTRWAVLVAGSNYYNYRHQAIDCHAYQLRKGLKDENTIVPMYDDIAPSS	95
Db	49	ENDDDSNGSTRWAVLVAGSSGYNYRHQAIDCHAYQLRKGLKEENVPMYDDIANNY	108
Qy	96	ENPRPGVIINKPDGEDGVTYKGPKDYTKAEVNVQFNYYNLGNEGSGVTGGNKVKVGSPND	155
Db	109	ENPRPGVIINSKHGVQGVQPKDYTGDDVNVDNLFVAILGDKTAVKGGSKGVVDSPND	168
Qy	156	NIFTYADHGAPGLIAMPTGDTEVMKAOFNEVLEKMHRKKYNKNMVIYEAECESGMFEGI	215
Db	169	HIFTYSDHGGPGVLGMPSTPSLYLANLDNLDVKKHALGTYSKLVFLYLEACESGIPEGL	228
Qy	216	LKKMLNIYATTAANSKSSMGVCYPESPYPPESEIGTCLGDTFSISWLEDSDLHDMKET	275
Db	229	LPEGILNIYATTASNAEISSMCTYCPEEPSPPEPYETCLGDLYSVANMEDSGMINLOTET	288
Qy	276	LEQQYHYVVKRR---VGSDVPETSHVCRFGTECKMLDULSSYIGNPENDNFTFE--SFS	330
Db	289	LHQOYELVKRRTAPVGSY--GSHVMQYGDVGISKDNLDLYMGTPNPANDNFTFADANSLK	346
Qy	331	SPISNSGLVNPDRDIPPLYLORIKOKAPWSLESKAOKLLDEKNIRKOIDOSTITDLRL	390
Db	347	PP---SRVTNORDADLVHFWEKYRKAPEGSAARKTEAQKVLEAMSHRLHIDSVLYGKI	403
Qy	391	SVKOTNVNLITSTRITTCGOLPVDDDCFKTLVNSFKNHCGATVHYGLTYCALANICNMG	450
Db	404	LFGISRGPVELVKNRSAGOPPLVDNNCLKNOVFARERHCSSLGSOYGINHMRSFANICNAG	463
Qy	451	VDVKQTVSATEQACS	465
Db	464	IOMEERAAASQACT	478

RESULT 4
TI2043
probable legumin (EC 3.4.22.34) precursor - kidney bean
N:Alternate names: asparaginyl endopeptidase; bean endopeptidase; phaseollin; vicillin
C:Species: Phaseolus vulgaris (kidney bean)
C:date: 16-Jul-1999 #sequence_revision 16-Jun-1999 #text_change 21-Jul-2000
C:Accession: TI2043
R:Senuyk, V.; Rotari, V.; Becker, C.; Zaharou, A.; Muentz, K.; Vaintraub, I.; Horstman
Eur. J. Biochem. 258, 546-558, 1998
A:title: Does an asparaginyl-specific cysteine endopeptidase trigger phaseolin degrad
A:Reference number: Z17389; Mjuid:99089618; PMID:9874222
A:Accession: TI2043
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-484 <SEN>
A:Cross-references: EMBL:Z99956; NID:q2511696; PIDN:CABL7078.1; PID:q2511697

RESULT 3
T05302
vacuolar processing enzyme (EC 3.4.22.-) isozyme gamma precursor - Arabidopsis thaliana
N;Alternate names: protein F26p21.60
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 09-Jun-2000
C;Accession: T05302

A: Note: the authors translated the codon TTC for residue 135 as Leu and the codon TAC for

C: Genetics:

A: Introns: 68/3; 123/3; 176/1; 204/3; 271/2; 287/3; 355/3; 422/3

C: Superfamily: legumain

C: Superfamily: legumain

C: Keywords: cysteine proteinase; glycoprotein; hydrolase

F: 12-46/Domain: amino-terminal propeptide #status predicted <PRO>

F: 47-46/Domain: vacuolar processing enzyme isozyme beta #status predicted <PRO>

F: 362-484/Domain: carboxyl-terminal propeptide #status predicted <PRO>

F: 307/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 46.8%; Score 1148; DB 2; Length 484;

Best Local Similarity 48.7%; Pred. No. 1.2e-71;

Matches 228; Conservative 69; Mismatches 153; Indels 18; Gaps 5;

QY 7 HFQILVFLHALLIFSASR-----KTQLNDNDVSSDSKAGTRWAVLVAGSNEY 57

DB 6 YFRPALLLLLVHAEGRPEPKILMPTTEANPAD---QDEGCVGRWAVLVAGSSGY 62

QY 58 YNRHQADICHAYQILRKGLKDENIIVPMYDDIAFSSSENPRPGVIINKPGEDVYKGV 117

DB 63 GNYRHQADVCHAYQILRKGLKEENIVLMYDDIANHPLNRPRTGLINHPDGDVYAGVP 122

QY 118 KYTKEAVNVQNFYVLLGNESGVTGGNGKVKVKGSPNDNFIYVADHAGPLIAMPTGDE 177

DB 123 KYTGSVTAANFYVLLGDKAVKGGCKVIASKPNDFIYVADHAGPLIAMPTPH 182

QY 178 VMAKDENEVLEKMKRKKYKVMYVVEACSGSMFEGILKKNLYAVTAANSKSSGV 237

DB 183 IYAADFIEFLKKAHSGTYKEMVYVVEACSGSIFEGIMPKDLNIYVTTASNAQESSYGT 242

QY 238 YCPESYPPPESETGTCGLDFTSLEWDSLDHMSKETLEQYHVVKRVGSDVPETSHV 297

DB 243 YCPGMPSPPESEYITCLDGLSVANWEDSETHNLKKEIKQYHVMRTSNYNTYSGSHV 302

QY 298 CRGTGEMKLDYLSYIGNPNENFTFESFSSISGLVNPDPILLYLQRIQAP 357

DB 303 MEYNSIKSEKILYQGFEPATVNLPLNELPAK--SKIGVNVQDADLLFLHMHRTSE 360

QY 358 MGSLESKEAKLLDEKHKRQIDOSTIDILRLSVKQT--NVNLLTSTRTTGQPLVDWD 416

DB 361 DGRKKDDTLKELTETTRHKHLDASVELIATILFGPTMNVNLV---REPGLPLVDWME 417

QY 417 CFKTLVSNFKHCGATVHYGLKTYGALANICNMGVQDVQVSAIEQAC 464

DB 418 CLKSMYRVFEEHCGSLTQYGMKHMRAFANVCNNGVSKELMEEAATAAC 465

RESULT 7

C96652

protein F23N19.7 [imported] - Arabidopsis thaliana

C: Species: Arabidopsis thaliana (mouse-ear cress)

C: Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 23-Mar-2001

C: Accession: C96652

R: Theologos, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

Chen, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;

ansen, N.F.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A: Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A: Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,

ker, W.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A: Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A: Reference number: A86141; MUID: 21016719; PMID: 11130712

A: Accession: C96652

A: Status: preliminary

A: Molecule type: DNA

A: Residues: 1-536 <STO>

A: Cross-references: GB:AE005173; NID: g6630462; PIDN: AAF19550.1; GSPDB: GN00141

C: Genetics:

A: Gene: F23N19.7

A: Map position: 1

C: Superfamily: legumain

Query Match 45.8%; Score 1125; DB 2; Length 536;

Best Local Similarity 44.5%; Pred. No. 5.5e-70;

Matches 232; Conservative 68; Mismatches 149; Indels 72; Gaps 8;

QY 7 HFQILVFLHALLIFSASR-----KTQLNDNDVSSDSKAGTRWAVLVAGSNEY 57

DB 6 YFRPALLLLLVHAEGRPEPKILMPTTEANPAD---QDEGCVGRWAVLVAGSSGY 62

QY 58 YNRHQADICHAYQILRKGLKDENIIVPMYDDIAFSSSENPRPGVIINKPGEDVYKGV 117

DB 63 GNYRHQADVCHAYQILRKGLKEENIVLMYDDIANHPLNRPRTGLINHPDGDVYAGVP 122

QY 118 K-----DYTKEAVNV 127

DB 123 KALHNYSDSDCRDICVCKPNEFCGPIGAPFLIATICSIVYLYKLYFODYTGSVTA 182

QY 128 QNFYVLLGNESGVTGGNGKVKVKGSPNDNFIYVADHAGPLIAMPTGDEVMAKDFNEVL 187

DB 183 ANFYAVLLGDKAVKGGCKVIASKPNDFIYVADHAGPLIAMPTPHIYAADFIEFL 242

QY 188 EKMKRKKYKVMYVVEACSGSMFEGILKKNLYAVTAANSKSSGVYCPESYPPPP 247

DB 243 KKHASCTYKEMVYVVEACSGSIFEGIMPKDLNIYVTTASNAQESSYGTCPGMNPP 302

QY 248 SEITGCTLGDTFSLEWDSLDHMSKETLEQYHVVKRVG--SDVPETSHVCRFGTEKM 305

DB 303 SEYITCLDGLSVANWEDSETHNLKKEIKQYHVMRTSNYNTYSGSHVMEYGNNSI 362

QY 306 LKDYLSYIGRNPENFTFESFSSPL--SNGLVNPDPILLYLQRIQAKPMGSLSK 364

DB 363 KSEKLYIQGFDPATVNLPLNE---LPVASKIGVNVQDADLLFLHMHYRTSEDSRKKD 419

QY 365 EAOKLLDEKHKRQIDOSTIDILRLSVKQT--NVNLLTSTRTTGQPLVDWDCTKTLVN 423

DB 420 DTLEKLTETTRHKHLDASVELIATILFGPTMNVNLV---REPGLPLVDWDECLKSMVR 476

QY 424 SFKNHCGATVHYGLKTYGALANICNMGVQDVQVSAIEQAC 464

DB 477 VFEHCGSLTQYGMKHMRAFANVCNNGVSKELMEEAATAAC 517

RESULT 8

T12044

probable legumain (EC 3.4.22.34) precursor - kidney bean

N: Alternate names: asparaginyl endopeptidase; bean endopeptidase; phaseolin; vicilin

C: Species: Phaseolus vulgaris (kidney bean)

C: Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 20-Jun-2000

C: Accession: T12044

R: Senyuk, V.; Becker, C.; Muentz, K.

submitted to the EMBL Data Library, October 1997

A: Description: Isolation of cDNA clone encoding legumain-like proteinase (LLP2) from

A: Reference number: 217390

A: Accession: T12044

A: Status: preliminary; translated from GB/EMBL/DBDJ

A: Molecule type: DNA

A: Residues: 1-493 <SEN>

A: Cross-references: EMBL: Z99957

A: Experimental source: cultivar Moldavian; cotyledon; clone p21b

C: Superfamily: legumain

C: Keywords: cysteine proteinase; hydrolase

F: 1-53/Domain: propeptide #status predicted <PRO>

F: 54-493/Product: probable legumain #status predicted <MAT>

Query Match 45.5%; Score 1115.5; DB 2; Length 493;

Best Local Similarity 50.1%; Pred. No. 2.2e-69;

Matches 220; Conservative 68; Mismatches 128; Indels 23; Gaps 5;

QY 34 DVSSDSKAGTRWAVLVAGSNEYNYRHQADICHAYQILRKGLKDENIIVPMYDDIAF 93

DB 49 DAESDE---VGRWAVLVAGSNGYNYRHQADYCHAYQILKGVKEENIVVPMYDDIAT 105

QY 94 SSENPRGVIINPDGDDVYGVKPKDYTKAVNVQNFYNNVLLGNESGVTGGNGKVKVSGP 153
 Db 106 HELNPRGVIINPDGDDVYGVKPKDYTKAVNVQNFYNNVLLGNESGVTGGNGKVKVSGP 165
 QY 154 NDNIYIYADHAGCLTAMPTGDEVMADKDFNEVLEKMKHKKYKWKWYIYVEACSGSWFE 213
 Db 166 EDRIYVYSDHGGVGLGMPNMPYLYAMDFIDVLKKHAGSGYKEMWYIYVEACSGSIFE 225
 QY 214 GILKKNLIYAVTAANSKESGWYVCPSPPPSEIGTCLGDTFISWLEDSDLHDSMK 273
 Db 226 GIMPKDLNIYVTTASNAQESNGTYCGMPPPPPEYITCLGDLYSVAMWEDSESHNLK 285
 QY 274 ETLEQQYHVVKRRVGS--DVPETSHVCRFGTEKMLDKYLSYIGRNPENDNFTTFESFS 331
 Db 286 ESVEQQYQSVKQRTSNFEAVAMGSHVMQYGDANMTAEKLYLHGFDPATVNF----- 337
 QY 332 PISNSGL-----VNPDIPLLYLQRIKAPMGSLSKAEQKLLDEKNHRKQIDOSIT 385
 Db 338 PPHNGRLSKMEVYNQDAELHFMWOLYKRSSENGSEKKKEILQOIKDAIKHRSHLDSSMQL 397
 QY 386 DILRLSVKQTNVLLNSTRTTGOPLVDDMCFKTLVNSFKNHCATVHYGLKYTGALAN 445
 Db 398 LIGVLLYCPKASSVLRVITGLPLVDWDTCLASWVRVYETHCGSLTQYGMKHKHRAFAN 457
 QY 446 ICNMGVDVQTVSAIEQAC 464
 Db 458 ICNSGV-----SETSMERAC 472

RESULT 9
 JQ2387
 vacuolar processing enzyme (EC 3.4.22.-) precursor - castor bean
 C:Species: Ricinus communis (castor bean)
 C:Date: 28-Aug-1985 #sequence_revision 07-Oct-1994 #text_change 20-Jun-2000
 C:Accession: JQ2387; PQ0863
 R:Hara-Nishimura, I.; Takeuchi, Y.; Nishimura, M.
 Plant Cell 5, 1651-1659, 1993
 A:Title: Molecular characterization of a vacuolar processing enzyme related to a putative
 A:Reference number: JQ2387; MUID:94146557; PMID:8312744
 A:Accession: JQ2387
 A:Molecule type: mRNA
 A:Residues: 1-497 <HAR>
 A:Cross-references: GB:DL7401; NID:g471161; PIDN:BA04225.1; PID:g471162
 A:Accession: PQ0863
 A:Molecule type: protein
 A:Residues: 104-141 <HA2>
 A:Experimental source: seed
 C:Comment: This enzyme plays a crucial role in the biosynthesis of vacuolar components a
 C:Superfamily: legumain
 C:Keywords: cysteine proteinase; glycoprotein; hydrolase; pyroglutamic acid
 F:1-31/Domain: signal sequence #status predicted <SIG>
 F:32-57/Domain: propeptide #status predicted <PRO>
 F:58-497/Product: vacuolar processing enzyme #status predicted <MAT1>
 F:58-374/Product: 37k vacuolar processing enzyme #status predicted <MAT2>
 F:375-497/Domain: carboxyl-terminal propeptide #status predicted <CTP>
 F:58/Modified site: pyrrolyl-terminal carboxylic acid (Gln) (in mature form)
 F:320/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 45.2%; Score 1109; DB 2; Length 497;
 Best Local Similarity 50.1%; Pred. No. 6.3e-69;
 Matches 220; Conservative 62; Mismatches 141; Indels 16; Gaps 3;

QY 35 VESSDKAKGTTRWAVLVAGSNEYNYRHQADICHAYQILRKGLKDNIIYFVYDDIAFS 94
 Db 51 VQVDDDLQGLTRWAVLVAGSNGFCNRYHQADVCHAYQILRKGLKDNIIYFVYDDIAFS 110
 QY 95 SENPRPGVIINPDGDDVYGVKPKDYTKAVNVQNFYNNVLLGNESGVTGGNGKVKVSGPN 154
 Db 111 ELNPRPGVIINPDGDDVYGVKPKDYTKAVNVQNFYNNVLLGNESGVTGGNGKVKVSGPN 170
 QY 155 DNIFIYADHAGCLTAMPTGDEVMADKDFNEVLEKMKHKKYKWKWYIYVEACSGSWFEG 214
 Db 171 DRIFLYYSDHGGVGLGMPNMPYLYAMDFIDVLKKHAGSGYKEMWYIYVEACSGSIFE 230

QY 215 ILKKNLIYAVTAANSKESGWYVCPSPPPSEIGTCLGDTFISWLEDSDLHDSMK 274
 Db 231 IMPKDVIIYVTTASNAQESNGTYCGMPPPPPEYITCLGDLYSVAMWEDSESHNLK 290
 QY 275 TLEQQYHVVKRRVGS--SDVPETSHVCRFGTEKMLDKYLSYIGRNPENDNFTTFESFSSP 332
 Db 291 TVKQYSSVARTSNYNTYAGSHVMQYGNQSIKADKLYLHGFDPASVNF-----P 342
 QY 333 ISNSGL-----VNPDIPLLYLQRIKAPMGSLSKAEQKLLDEKNHRKQIDOSITD 386
 Db 343 PNAHLNAPMEVYNQDAELHFMWOLYKRSSENGSEKKKEILQOIKDAIKHRSHLDSSMQL 402
 QY 387 ILRLSVKQTNVLLNSTRTTGOPLVDDMCFKTLVNSFKNHCATVHYGLKYTGALANI 446
 Db 403 IGDLLFGPKKASATLKSVPSPPLVDWDTCLASWVRVYETHCGSLTQYGMKHKHRTANI 462
 QY 447 CNMGVDVQTVSAIEQACS 465
 Db 463 CNAGVSHTSMEACNAACS 481

RESULT 10
 JX0344
 legumain (EC 3.4.22.34) precursor - jack bean
 N:Alternate names: asparaginyl endopeptidase
 C:Species: Canavalia ensiformis (jack bean)
 C:Date: 22-Apr-1995 #sequence_revision 26-May-1995 #text_change 21-Jul-2000
 C:Accession: JX0344
 R:Takeda, O.; Miura, Y.; Mita, M.; Matsushita, H.; Kato, I.; Abe, Y.; Yokosawa, H.;
 J. Biochem. 116, 541-546, 1994
 A:Title: Isolation and analysis of cDNA encoding a precursor of Canavalia ensiformis
 A:Reference number: JX0344; MUID:95155263; PMID:7852272
 A:Accession: JX0344
 A:Molecule type: mRNA
 A:Residues: 1-475 <TAK>
 A:Cross-references: DDBJ:D31787; NID:g499293; PIDN:BA06596.1; PID:g499294
 A:Experimental source: seed
 C:Comment: This enzyme is involved in posttranslational processing of concanavalin A
 C:Superfamily: legumain
 C:Keywords: cysteine proteinase; glycoprotein; hydrolase
 F:1-15/Domain: signal sequence #status predicted <SIG>
 F:16-35/Domain: propeptide #status predicted <PRO>
 F:36-475/Product: legumain #status experimental <MAT>
 F:300/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 44.0%; Score 1080; DB 2; Length 475;
 Best Local Similarity 48.3%; Pred. No. 5.9e-67;
 Matches 207; Conservative 73; Mismatches 129; Indels 20; Gaps 4;

QY 44 GTRWAVLVAGSNEYNYRHQADICHAYQILRKGLKDNIIYFVYDDIAFSSENPRPGVI 103
 Db 38 GTRWAVLVAGSNGVCNRYHQADVCHAYQILKGGKKNIIYFVYDDIAFNAMPRPGVI 97
 QY 104 INKPDGEDVYGVKPKDYTKAVNVQNFYNNVLLGNESGVTGGNGKVKVSGPNDFIYAD 163
 Db 98 INHPQPDVYAGVPKDYTGEDVTENLYAVILGDKSVKGVKINSNPEDRIFIFYSD 157
 QY 164 HGAPGLTAMPTGDEVMADKDFNEVLEKMKHKKYKWKWYIYVEACSGSWFEGILKKNLIY 223
 Db 158 HGGPGVGLGMPNAPVYAMDFIDVLKKHAGSGYKEMWYIYVEACSGSIFEIMPDKDLNIY 217
 QY 224 AVTAANSKESGWYVCPSPPPSEIGTCLGDTFISWLEDSDLHDSMKETLEQQYHV 283
 Db 218 VTTASNAQESNGTYCGMPPPPPEYITCLGDLYSVAMWEDSETHNLKRETVOQQQSV 277
 QY 284 KRRVGSVDPET--SHVCRFGTEKMLDKYLSYIGRNPENDNFTTFESFSSISNGL--- 338
 Db 278 RKRTSNNSYRFGSHVMQYGDNTITAEKLYLHGFDPATVNF-----PPHNGNLEAK 329
 QY 339 ---VNPDIPLLYLQRIKAPMGSLSKAEQKLLDEKNHRKQIDOSITDILRLSVKQT 395
 Db 330 MEVYNQDAELHFMWQYRSNHPQKTHILEQITETVHRNHLGDSVELLIGVLLYCPG 389

QY 396 NVLNLITSTRTTGOPLVDDHDCFKTLVNSFKNHCAGTVHYGLKYTGALANICNGVDVKQ 455
 Db :
 390 KSSSVLSVRAPGRLPLVDDWTCLKSMRVRFETHCGSLTQYGMKHHRFAGNVCSGV---- 445
 QY 456 TVSAIQAC 464
 Db :
 446 SKASMEEAC 454

RESULT 11
 T07132
 Cysteine proteinase (EC 3.4.22.-) precursor [similarity] - soybean
 C:Species: Glycine max (soybean)
 C:Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 21-Jul-2000
 C:Accession: T07132
 R:Shimada, T.; Hiraiwa, N.; Nishimura, M.; Hara-Nishimura, I.
 Plant Cell Physiol. 35, 713-718, 1994
 A:Title: Vacuolar processing enzyme of soybean that converts proproteins to the corresponding mature proteins
 A:Reference number: Z15942; MUID:94356350; PMID:8075902
 A:Accession: T07132
 A>Status: preliminary; translated from GB/EMBL/DDBJ
 A:Molecule type: mRNA
 A:Residues: 1-495 <SHI>
 A:Cross-references: EMBL:D28876; NID:g511937; PIDN:BAA06030.1; PID:g511938
 A:Experimental source: cotyledon
 C:Superfamily: legumin
 C:Keywords: cysteine proteinase; hydrolase
 F:I-55/Domain: propeptide #status predicted <PRO>
 F:56-495/Product: cysteine proteinase #status predicted <MAT>

Query Match	43.6%	Score 1071;	DB 2;	Length 495;
Best Local Similarity	48.3%;	Pred. No. 2.6e-66;		
Matches	210;	Conservative 70;	Mismatches 135;	Indels 20; Gaps 4;

QY 38 SDKSAGTRWAVLVAGSNFYNYRHOADICHAYOILKKGLKDENIIIVFYDDIAFSSEN 97
 Db :
 52 ADSDEVGTRWAVLVAGSNGYGRHOADVCHAYOLLIKGLKEENIVVFYDDIATNELN 111
 QY 98 PRGVIINKPDGEDVYKGVPDYTKTEAVNVONPYNVLLGNESGVTGGNGKVAKSGPNONI 157
 Db :
 112 PRGVIINHPEGEDLVAGVPDVTGDNVTTFENFAVLGDGSKLKGSGRGVINSKPEDRI 171
 QY 158 FIYYADHGAPGLIAMPTGDEVMAKDNEVLEKMKHKRYKNMVIYVEACSGSMFEGLIK 217
 Db :
 172 FIYSDHGGPGILGNMPPLYAMDFTDLVKLKHASGYKEMYIYVEACSGSVFEGIMP 231
 QY 218 KNLNIYAVTAANSKESGWGYCPESPYPPESEICTCLGDTFESISWLDSLDHDMSKETLE 277
 Db :
 232 KDLNIYVTTASNQAENSWGTYPCGMDPSPPEYITLCGLDLYSVAMMEDSEAHNLKRCSVK 291
 QY 278 QQYHVVKRRVG--SDVPETSHVCRFGTEKMLKDYLSYIGRNPENDNFTEFSFSSPIN 335
 Db :
 292 QQYKSVKQRTSNFNVMGSHVMQYGDNTITAEKLYLGDFPATYVF-----PPQN 343
 QY 336 SGL-----VNRPDIPLLYLQRIQAPMGSLSKAQKLLDEKHKRHKOIDSITDLR 389
 Db :
 344 GRLETKMVVNQDAELFLWMQYQRNSHQSENKTDLKQIAETVKRKHIIDGSVELIGV 403
 QY 390 LSVKQTINVLNLTSTRTTGOPLVDDWDCKPTLYNSPKNHCCGATVHYGLKYTGALANICNM 449
 Db :
 404 LLXGPCGSGSVLSVRAPGSSLVDDMTCLKSMRVRFETHCGTLTQYGMKHHRFAGNVCSGV 463
 QY 450 GVDVKOTVSIEQAC 464
 Db :
 464 GV-----SEASMEEAC 474

RESULT 12
 TI0944
 Cysteine proteinase (EC 3.4.22.-) precursor - spring vetch
 C:Species: Vicia sativa (spring vetch, tare)
 C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 20-Jun-2000

C:Accession: T10944
 C:Fischer, J.; Becker, C.; Hillmer, S.; Horstmann, C.; Neubohn, B.; Senyuk, V.; Muent
 submitted to the EMBL data Library, August 1998
 A:Description: The family of cysteine proteinases from *Vicia* seeds.
 A:Reference number: Z17221
 A:Accession: T10944
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-503 <PIS>
 A:Cross-references: EMBL:AJ007743
 A:Experimental source: cotyledons; clone pHS1
 C:Function:
 A:Description: plays a crucial role in the biosynthesis of vacuolar components and re
 C:Superfamily: legumin
 C:Keywords: cysteine proteinase; glycoprotein; hydrolase
 F:1-38/Domain: signal sequence #status predicted <SIG>
 F:39-62/Domain: propeptide #status predicted <PRO>
 F:63-503/Product: cysteine proteinase #status predicted <MAT>
 F:328/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 43.0%; Score 1054; DB 2; Length 503;
 Best Local Similarity 45.2%; Pred. No. 4e-65;
 Matches 213; Conservative 82; Mismatches 146; Indels 30; Gaps 7;

QY	10	ILVFLHALLIFSAESKLTOL-----LNDNDVESDSKSAKTRMAVLVAGSNQYNYRH	62
DB	26	LLFLSL--HGSVARNRLEWEPVRLDCEPDADVEIGTRMAVLVAGSNQYNYRH	83
QY	63	QADICHAYQILRRGGLKDENIIVFMDDDTAFSSENPRPGVIINKPDGEVYKGVKPDYTK	122
DB	84	QADVCHAYQLLIIGGVKEENIVFMDDDTAFSEPNRPGVIINHPGPNVYDGVKPDYTG	143
QY	123	EAVNQFVNYVLNGSGVYTGNGKVKVSGPNDNIFIYYADHGAAGLIAMPTGDEVMAKD	182
DB	144	DFTADNLXAVILGDSKVRGSGKVIINSKAEDRIPIYSDHGGPGVLGNPNPYVYAMD	203
QY	183	FNEVEKMIKKRKYNIYVEACESGMPGEGLKLNLIYAVTAANSKSSMGVY-CPE	241
DB	204	FIDVLKXKXASRGYQOMVIYVEACESGVFGQIMPKDIDVYVTTASNAEESNGTYLVPG	263
QY	242	SYPPPPSETGTCUGDTFTSLWLEDSDLHDMSKSTLQOQYHVVKRRV--GSDVPETSHVCR	299
DB	264	VYPASPEYITCUGDLYSVAMMDESETHNLKRETLKQOFASVKERTLNNNNYGLGSHVTE	323
QY	300	FGTEKMLKDYLSYVIGRNPENDNFTETESFSPISNSGL-----VNPRIPIPLYLQRKI	353
DB	324	YGTNTNTTDEKLYLHGFDPAFVNL-----PPNNGRLSKMEVYNQDAETLFWNQMY	375
QY	354	QKAPMGSLSEKQAKLLDEKNIRKOIDOSITDILRLSKQTNVLNLLTSTRITGTOPLVD	413
DB	376	QLRDLHQSEKKRDILKKISETVTKHRNLDGSEVELIGVLLFGPTRGSSVLQSVRASGLPLVD	435
QY	414	DWDGCFKTLVNSFKNHCAGTVHYGLKYTGALANICNMGVDVKOTVSALEQAC	464
DB	436	DWECLSKRVRFVETHGSLGSLTOYGMKHHRAFNALCNICNGI----SEDLMEETC	482

RESULT 13
 A60145
 C:Species: Schistosoma mansoni
 C:hemogloblinase (EC 3.4.-.-) precursor - fluke (Schistosoma mansoni)
 C:Date: 20-Feb-1993 #sequence_revision 20-Feb-1993 #text_change 09-Jun-2000
 C:Accession: A60145; B60145; A27378
 R:El. Manawy, M.A.; Ajli, T.; Phillips, N.F.B.; Davis, R.E.; Salata, R.A.; Malhotra, I
 Am. J. Trop. Med. Hyg. 43, 67-78, 1990
 A:Title: Definition of the complete schistosoma mansoni hemoglobinase mRNA sequence a
 A:Reference number: A60145; MUID:90342941; PMID:2382765
 A:Accession: A60145
 A:Status: not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 1-429 <ELA>
 A:Accession: B60145
 A:Molecule type: protein

A:Residues: 32-57 <EL2>
R:Davis, A.H.; Nanduri, J.; Watson, D.C.
J. Biol. Chem. 262, 12851-12855, 1987
A:Title: Cloning and gene expression of Schistosoma mansoni protease.
A:Reference number: A27378; MUID:87308326; PMID:3305515
A:Accession: A27378
A:Molecule type: mRNA
A:Residues: 77-309, 'V', 311-429 <DAV>
A:Cross-references: GB:M17423; NID:g161060; PID:g161061
C:Superfamily: legumain
C:Keywords: glycoprotein; hydrolase; proteinase
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-31/Domain: propeptide #status predicted <PRO>
F:32-291/Product: hemoglobinase #status predicted <MAT>
F:292-429/Domain: carboxyl-terminal propeptide #status predicted <Cmp>
F:192,205,215/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 32.2%; Score 790; DB 2; Length 429;
Best Local Similarity 40.5%; Pred. No. 5.6e-47;
Matches 186; Conservative 75; Mismatches 156; Indels 42; Gaps 13;

QY 10 ILVFLHALLIPSAESRKQTLNDNDVESSDKS-AKGTWAVLVAGSNFYNYRHOADICH 68
DB 8 LISILHLLV-----KQCL--DTYEVSDFTVSDNNKAVLVAGSNGTPNTRHQADVCH 59
QY 69 AVQILRKGLKNDENIIVFYDDIAPSSNPGRGVIINPKDGVYKGVPKDYTKAVNVQ 128
DB 60 AVHLVRSKGIRPEHIITWYDDIAVNLNPPPKLFNDYNHKKDWYEGVVYDYGKRVNSK 119
QY 129 NFYNVLLGNESVTCGNGKVKVSGNDNFIYADHAGPLIAMPDTGDEVMAKDFNEVLE 188
DB 120 TFLKVLKGDKS----AGGKVLKSGKNDVFIYFDHAGPLIAFP--DDELVAKEFMSTLK 174
QY 189 KHKRKKYKMKVIYVEACSGSMFEGILKLNLIYAVTAANSKSSWGVCYCPESYPPPPS 248
DB 175 YLHSHKRYSKLVIYIYANESGSMFOILPSNLISVATTAANSSTCYSTFCGD-----P 228
QY 249 EIGTCLGDTFTSWLESDLDHMSKETLEQQYHVYKRRVGVSDVPETSHVCRFGTEKMLKD 308
DB 229 TITTCCLADLYSNVIVDSQTHLTQRTLDQYKVKRET-----DLSHVQRYGDTRMGKL 283
QY 309 YLSSYIGRNPENDNTFTFESSPSNSGLNPNRDIPLLYLQKTKAPMGSLSKSAQK 368
DB 284 YVSEFOGSRDKS-----SENDEPPMKPRHSIASRDIPLTLHROIMT--NNAEDKSFLLM 337
QY 369 KLLDEKNRKKQIDQSIDTLRLSVKQTNVNLTLSTRTTGOPLVDWDCFKTLVNSEKNH 428
DB 338 QILGLKLRRDL---IEDTMKLIIVKMNEEIPNTKATIDQTL----DCTESVYEQFKSK 390
QY 429 CGATVHYGLKYTG----ALANICNMGVGVKQTVSAIEQAC 464
DB 391 C-FTLQQAPEVGGHFSTLYNCADGYTAETINEAIIKIC 428

RESULT 14
T19231
probable cysteine proteinase (EC 3.4.22.-) T28H10.3, precursor [similarity] - Caenorhabditis
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
C:Accession: T19231; T25439
R:Dobson, R.
submitted to the EMBL Data Library, July 1996
A:Reference number: 219094
A:Accession: T19231
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-462 <WIL>
A:Cross-references: EMBL:Z77653; PIDN:CA801126.1; GSPDB:GN00023; CESP:T28H10.3
A:Experimental source: clone C13C12
R:Kershaw, J.
submitted to the EMBL Data Library, June 1996
A:Reference number: 220034
A:Accession: T25439

A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-462 <WIL>
A:Cross-references: EMBL:Z75551; PIDN:CAA99935.1; GSPDB:GN00023; CESP:T28H10.3
A:Experimental source: clone T28H10
C:Genetics:
A:Gene: CESP:T28H10.3
A:Map position: 5
A:Introns: 93/2; 148/2; 161/1; 218/1; 311/3; 379/2
C:Superfamily: legumain
C:Keywords: cysteine proteinase; glycoprotein; hydrolase
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-39/Domain: propeptide #status predicted <PRO>
F:40-462/Product: probable cysteine proteinase T28H10.3 #status predicted <MAT>
F:134,278/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 31.9%; Score 782.5; DB 2; Length 462;
Best Local Similarity 38.8%; Pred. No. 2.1e-46;
Matches 169; Conservative 81; Mismatches 145; Indels 41; Gaps 9;

QY 43 KGTRWAVLVAGSNFYNYRHOADICHAYOILRKGLKNDENIIVFYDDIAPSSNPGRGV 102
DB 40 EGEAFVVLVAGSNGWYNYRHOADVAHAHTLRNHGIPENIITMYDDVANNPLNPKYK 99
QY 103 IINKPDGEDVYKGVPKDYTKAVNVQNFYNYVLLGNESVTCGNGKVKVSGPNDNFIY 162
DB 100 LFNRPBGKDYKGLIDYKGVASVTPENFLNLKGNASGIDGGNGRVLETNDNDRVFYFT 159
QY 163 DHGAPGLIAMPDTGDEVMAKDFNEVLEKHKRKKYKMKVIYVEACSGSMFEGILKLNLI 222
DB 160 DHGAVCMISFPDG--ILTQKQNDVLVMMHKNKYSQLTFLYLEACSGSMFEVLRSMDI 218
QY 223 YAVTAANSKSSWGVCYCPESYPPPPSEIGTCLGDTFTSWLESDLDHMSKETLEQQYHV 282
DB 219 YAISSAASHSSWGTCFCDNMLP-----CLGDLFSVNNWTDSDGDKLTLEFQVEL 272
QY 283 VKRRVGVSDVPETSHVCRFGTEKMLKDYLSYIGRNPENDNTFTFESSPSNSGLVNP 342
DB 273 VKKET-----NLSHVMQFGDKDIAKEAVALFOG---DKEDREYVEDFGLSASKSVNMPAR 324
QY 343 DIPLLYLQKTKAPMGSLSKSAQKLLDEKNHKKIDOSITDILRL---SVKQTNVLN 399
DB 325 DIELNHLISQHRKS--NDLLSSNKLEYKINRIKETRAIKRNVHMIVQKFFDGESEDLISR 383
QY 400 LITSRTTGOPLVDWDCFKTLVNSEKNHC--GATVHYGLKYTGALANIC----- 447
DB 384 VLTQTRP-----VLDLRCHHTAVHLFKKYCINFEYEMKVKVYKVINNCIYRTEIEIVL 438
QY 448 -----NMGVGVKQTVS 458
DB 439 ALPDICMDIDIEQVA 454

RESULT 15
S31908
hemoglobinase - fluke (Schistosoma japonicum)
C:Species: Schistosoma japonicum
C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 09-Jun-2000
C:Accession: S31908
R:Merckelbach, A.; Hasse, S.; Dell, R.; Eschibeck, A.; Ruppel, A.
submitted to the EMBL Data Library, February 1993
A:Description: cDNA sequences of Schistosoma japonicum coding for hemoglobinase and t
A:Reference number: S31907
A:Accession: S31908
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-423 <MER>
A:Cross-references: EMBL:X70967; NID:g11164; PID:g11165
C:Superfamily: legumain

Query Match 31.3%; Score 768; DB 2; Length 423;
Best Local Similarity 38.4%; Pred. No. 1.8e-45;
Matches 178; Conservative 70; Mismatches 167; Indels 48; Gaps 13;

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 27, 2003, 15:12:13 ; Search time 14 Seconds
(without alignments)
1380.569 Million cell updates/sec

Title: US-09-934-066-2

Perfect score: 2454

Sequence: 1 MSSPLGHFQILVFLHLLIF.....CNGMGVDVKQTVSATEQACSM 456

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1268.5	51.7	493	1 VPE_VICSA	P49044 vicia sativ
2	1262.5	51.4	494	1 VPE_CITSI.	P49043 citrus sine
3	1246	50.8	490	1 VPEG_ARATH	Q39119 arabidopsis
4	1230.5	50.1	478	1 VPEA_ARATH	P49047 arabidopsis
5	1160	47.3	486	1 VPEB_ARATH	Q39044 arabidopsis
6	1109	45.2	497	1 VPEB_RICCO	P49042 ricinus com
7	1080	44.0	475	1 LEGU_CANEN	P49046 canavalia e
8	1071	43.6	495	1 VPE_SOYBN	P49045 glycine max
9	843	34.4	435	1 LGNN_MOUSE	O89017 mus musculus
10	832	33.9	435	1 LGNN_RAT	Q9R0J8 rattus norv
11	828	33.7	433	1 LGNN_HUMAN	Q99538 homo sapien
12	790	32.2	429	1 HGLB_SCHWA	P09841 schistosoma
13	768	31.3	423	1 HGLB_SCHJA	P42665 schistosoma
14	321	13.1	411	1 GP18_YEAST	P49018 saccharomyc
15	297.5	12.1	395	1 GP18_HUMAN	Q92643 homo sapien
16	292.5	11.9	380	1 GP18_SCHPO	Q99535 schizosacch
17	280.5	11.4	395	1 GP18_MOUSE	Q9CXY9 mus musculus
18	276	11.2	322	1 GP18_CAEEL	P49048 caenorhabdi
19	116	4.7	522	1 IBMP_CAMVB	P16666 cauliflowe
20	115	4.7	872	1 DPOL_SULOH	O50607 sulfurispha
21	114.5	4.7	580	1 DAK1_SCHPO	OL3902 schizosacch
22	112.5	4.6	1440	1 STEP_HUMAN	P07814 homo sapien
23	111.5	4.5	1164	1 KEL1_YEAST	P39853 saccharomyc
24	109	4.4	1714	1 ITN1_MOUSE	Q920R4 mus musculus
25	106.5	4.3	1059	1 XYN1_THEMA	O60037 thermotoga
26	105.5	4.3	800	1 PTL1_YEAST	P08468 saccharomyc
27	104.5	4.3	369	1 LEU2_BUCUL	Q9AQC6 buchnera ap
28	104.5	4.3	451	1 GP1D_CHLMU	P06437 chlamydia m
29	104	4.2	442	1 PUS3_YEAST	P31115 saccharomyc
30	103.5	4.2	520	1 TIMP_ECOLI	Q47163 escherichia
31	103.5	4.2	755	1 P100_HSV7J	P52519 human herpe
32	103	4.2	417	1 YGX3_YEAST	P53079 saccharomyc
33	103	4.2	3175	1 RPOA_EAV	P19811 equine arte

RESULT 1

VPE_VICSA STANDARD; PRT: 493 AA.

AC P49044;

DT 01-FEB-1996 (Rel. 33, Created)

DT 01-FEB-1996 (Rel. 33, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Vacuolar processing enzyme precursor (EC 3.4.22.-) (VPE) (Proteinase B).

OS Vicia sativa (Spring vetch) (Tare).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eurosids I; Fabales; Fabaceae; Papilionoideae; Viciae; Vicia.

OX NCBI_TaxID=3908;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Seed;

RX MEDLINE=95220376; PubMed=7705362;

RA Fischer C., Shutov A.D., Nong V.H., Senyuk V.I., Jung R., Horstmann C.,

RA "Purification, cDNA cloning and characterization of proteinase B, an

RT asparagine-specific endopeptidase from germinating vetch (Vicia

RT sativa L.) seeds.";

RL Eur. J. Biochem. 228:456-462(1995).

RN [2]

RC CHARACTERIZATION.

TX MEDLINE=82232279; PubMed=7046813;

RA Shutov A.D., Do N.L., Vaintraub I.A.;

RT "Purification and partial characterization of protease B from

RT germinating vetch seeds.";

RL Biochimica 47:814-821(1982).

CC -!- FUNCTION: ASPARAGINE-SPECIFIC ENDOPEPTIDASE INVOLVED IN THE

CC PROCESSING OF VACUOLAR SEED PROTEIN PRECURSORS INTO THE MATURE

CC FORMS.

CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C13.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

CC EMBL: Z34899; CAA84383.1; -

DR MEROPS: C13.002; -

DR InterPro: IPR001096; Legumain.

DR Pfam: PF01650; Peptidase_C13; 1.

DR PRINTS: PRO0776; HEMOGLOBINASE.

KW Hydrolase; Thiol protease; Signal.

FT SIGNAL 1 19 POTENTIAL.

FT CHAIN 20 493 VACUOLAR PROCESSING ENZYME.

FT ACT_SITE 174 174 POTENTIAL.

FT ACT_SITE 216 216 POTENTIAL.

FT VARIANT 88 88 S -> L.

FT

```

FT VARIANT 89 228 228 K -> R.
FT VARIANT 228 228 D -> E.
FT VARIANT 254 254 P -> S.
FT VARIANT 366 366 R -> H.
SQ SEQUENCE 493 AA: 54383 MW: 54383 MW: FELFFBADAFFB63AE CRC64;

Query Match
Best Local Similarity 51.7%; Score 1268.5; DB 1; Length 493;
Matches 235; Conservative 70; Mismatches 122; Indels 11; Gaps 3;

QY 31 NDNVSSDKSAGTRWAVLVAGSNYYNRHQADICHAYQILRKGGGLKDNIIIVFMYD 90
DB 48 NDDDFE-----GTRWAILLAGSNYYNRHQSDVCHAYQILRKGGSKENIIIVFMYD 100

QY 91 IAFSSNPRPGVIINKPDGDDYVYKGVKDYKAEVNVQNFYVNLGNESGVTGGNGKVV 150
DB 101 IASNEENPRPGVIINKPDGDDYVYKGVKDYKAEVNVQNFYVNLGNESGVTGGNGKVV 160

QY 151 SGNPNDFIYYADHAGAPLIAMPTGDEVMAKDFNEVLKMKHKKYKMKWIVVEACESG 210
DB 161 SGNPNDFIYYADHAGAPLIAMPTGDEVMAKDFNEVLKMKHKKYKMKWIVVEACESG 220

QY 211 MFGILKKNLNIYAVTAANSKSSNGVYCPESYPPPEIGTCIGDTFSISWLESDILH 270
DB 221 IFGILLPDLNIIYATTASNAEESNGVYCPGDKPPPEYSTCLGLDLYSIAMMEDSEVHN 280

QY 271 MSKETLEQOYHVVKRRVSGSDVPETSHVCRFGTEKMLKDYLSYIGRNPENDNFTTESF 330
DB 281 LQTESLQOYKLVKNTISE-PYGRVWEIGDGLSKNDLYQLGTNPANDNNSFVDETE 339

QY 331 SPT---SNGSLVNPDRIPILLYLQRIQKAPMGSLKESKAEQKLLDEKNHRKQIDOSTDI 387
DB 340 NSLKLTPSAANORDADLTHFEWFRKRAPEGSSOKNAEAKQVLEAMSHRKHIDNSVKLI 399

QY 388 LRSVKQTNVNLTLSTRTTGOPLVDWDCFKTLVNSFKNHGATVHYGLKYTGALANIC 447
DB 400 GQLFGIEKTELLDVVRPAGSPLVDNWDCLTKMTYKTFETHCGLSQYGMKMRHSFANIC 459

QY 448 NMGVDVKQTVSAIEQACS 465
DB 460 NAGIPNEPMAEASQAAC 477

RESULT 2
VPE_CITSI STANDARD; PRT: 494 AA.
AC P49043;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Vacuolar processing enzyme precursor (EC 3.4.22.-) (VPE).
OS Citrus sinensis (Sweet orange).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC Ericoidae; Sapindales; Rutaceae; Citrus.
OX NCBI_taxid=2711;
RN SEQUENCE FROM N.A.
RC STRAIN=cv.washington Navel; TISSUE=Flavored;
RA MEDLINE=96030252; PubMed=7480346;
RA Alonso J.M., Granel A.;
RT also during fruit ripening in Citrus fruit."
RL plant Physiol. 109:541-547(1995).
CC -1- FUNCTION: ASPARAGINE-SPECIFIC ENDOPEPTIDASE THAT MAY BE INVOLVED
CC IN PROCESSING OF PROTEINS TARGETED TO VACUOLES THAT ACCUMULATE
CC DURING ETHYLENE-REGULATED PROCESSES SUCH AS FLOWER OPENING AND
CC FLAVOIDS DEGRADING.
CC -1- TISSUE SPECIFICITY: HIGH LEVELS ARE SEEN IN THE FLOWERS, A LOWER
CC LEVEL EXPRESSION IS SEEN IN THE LEAVES, WHILE VERY LOW LEVELS ARE
CC SEEN IN THE STEMS AND ROOTS.
CC -1- DEVELOPMENTAL STAGE: THE LEVELS ARE LOW IN GREEN FRUITS BUT
CC ACCUMULATE WITH COLOR CHANGE OCCURRING DURING RIPENING, REACHING

```

```

CC CC MAXIMUM LEVELS IN FULLY COLORED FRUIT. THE LEVELS INCREASE DURING
CC FLOWER DEVELOPMENT AND SHOW HIGHEST LEVELS IN FLOWERS AT ANTHESIS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C13
CC -----
CC CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC CC EMBL: Z47793; CAA87720.1; -
CC MEROPS: C13.002; -
CC DR InterPro: IPR001096; Legumain.
CC DR Pfam: PF01650; Peptidase_C13; 1.
CC DR PRINTS: PR00776; HEMOCLOENASE.
CC KW Hydrolyase; Thiol protease; Glycoprotein; Signal.
CC FT SIGNAL 1 20 POTENTIAL.
CC FT CHAIN 21 494 VACUOLAR PROCESSING ENZYME.
CC FT CARBOHYD 151 151 N-LINKED (GLCNAC. .) (POTENTIAL).
CC FT CARBOHYD 336 336 N-LINKED (GLCNAC. .) (POTENTIAL).
CC FT ACT_SITE 178 178 POTENTIAL.
CC FT ACT_SITE 220 220 POTENTIAL.
CC SQ SEQUENCE 494 AA: 54291 MW: 82059237E03B2B8.CRC64;

Query Match
Best Local Similarity 51.4%; Score 1262.5; DB 1; Length 494;
Matches 238; Conservative 69; Mismatches 120; Indels 9; Gaps 3;

QY 30 LNDNVSSDKSAGTRWAVLVAGSNYYNRHQADICHAYQILRKGGGLKDNIIIVFMYD 89
DB 50 VNDDE-----DSVGRVWAVLLAGSNGFWNRHQADICHAYQILRKGGGLKDNIIIVFMYD 103

QY 90 DIAFSSNPRPGVIINKPDGDDYVYKGVKDYKAEVNVQNFYVNLGNESGVTGGNGKVV 149
DB 104 DIAFNEENPRPGVIINKPDGDDYVYKGVKDYKAEVNVQNFYVNLGNESGVTGGNGKVV 163

QY 150 KSGPNDFIYYADHAGAPLIAMPTGDEVMAKDFNEVLKMKHKKYKMKWIVVEACESG 209
DB 164 DSGPNDFIYYADHAGAPLIAMPTGDEVMAKDFNEVLKMKHKKYKMKWIVVEACESG 223

QY 210 SMFEGILKKNLNIYAVTAANSKSSNGVYCPESYPPPEIGTCIGDTFSISWLESDILH 269
DB 224 SIFEGILLLEGLNIIYATTASNAEESNGVYCPGEPPEYSTCLGLDLYSIAMMEDSDIH 283

QY 270 DMSKETLEQOYHVVKRRVSGSDVPETSHVCRFGTEKMLKDYLSYIGRNPENDNFTTESF 329
DB 284 NLRTETLHQOYELVKRTASYNYSYGVSHVQYGDIGLSKNNLFTYLGTPANDNFTFVDEN 343

QY 330 S-SPISSNGLVNPDRIPILLYLQRIQKAPMGSLKESKAEQKLLDEKNHRKQIDOSTDIL 388
DB 344 SLRPASKA--VNORDADLHFWDKYRRAPEGTPRKAEGQFEAMSHRHWVHDSIKLIG 401

QY 389 RLSVKQTNVNLTLSTRTTGOPLVDWDCFKTLVNSFKNHGATVHYGLKYTGALANIC 448
DB 402 KLLFGIEKGEPEILNTVRPAGSPLVDNWDCLTKMTYKTFETHCGLSQYGMKMRHSFANIC 461

QY 449 MGVVDVKQTVSAIEQAC 464
DB 462 TGIGKEKMAEASQAAC 477

RESULT 3
VPE_ARATH STANDARD; PRT: 490 AA.
ID VPE_ARATH
AC Q39119; O95B73;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Vacuolar processing enzyme, gamma-isozyme precursor (EC 3.4.22.-)
DE (Gamma-VPE).
DE AT4G32940 OR F26P21.60.
GN

```


OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RC SEQUENCE FROM N.A.
 RP STRAIN=cv. Columbia;
 RC MEDLINE=96146075; PubMed=8589932;
 RX Kinoshita T., Nishimura M., Hara-Nishimura I.;
 RA "The sequence and expression of the gamma-VPE gene, one member of a
 RT family of three genes for vacuolar processing enzymes in Arabidopsis
 RT thaliana.";
 RL Plant Cell Physiol. 36:1555-1562(1995).
 RN [2]
 RC SEQUENCE FROM N.A.
 RP STRAIN=cv. Columbia;
 RC MEDLINE=20083488; PubMed=10617198;
 RX Mayer K.F.X., Schueller C., Wambutt R., Murphy G., Volckaert G.,
 RA Pohl T., Duesterhoeft A., Stiekema W., Entian K.-D., Terryn N.,
 RA Harris B., Ansoorge W., Brandt P., Grivell L.A., Rieger M.,
 RA Weichselgartner M., de Simone V., Obermaier B., Mache R., Mueller M.,
 RA Kreis M., Delseny M., Puigdomenech P., Watson M., Schmidheini T.,
 RA Reichert B., Portetelle D., Perez-Alonso M., Boutry M., Bancroft I.,
 RA Vos P., Hoheisel J., Zimmermann W., Wedler H., Ridley P.,
 RA Langham S.-A., McCullagh B., Bilham L., Robben J.,
 RA Van der Schueren J., Grymonprez B., Chuang Y.-J., Vandenbussche F.,
 RA Braeken M., Weltjens I., Voet M., Bastiaens I., Aert R., Defoor E.,
 RA Weltzenegger T., Bothe G., Ramsperger U., Hilbert H., Braun M.,
 RA Holzer E., Brandt A., Peters S., van Staveren M., Dirkse W.,
 RA Berneliser S., Hempel S., Feldpausch M., Lamberth S., van den Baele H.,
 RA De Keyser A., Buyschaert C., Gielen J., Villarroel R., De Clercq R.,
 RA Van Montagu M., Rogers J., Cronin A., Quail M., Bray-Allen S.,
 RA Clark L., Doggett J., Hall S., Kay M., Lennard N., McLay K., Mayes R.,
 RA Pettett A., Rajandream M.A., Lyne M., Benes V., Rechmann S.,
 RA Borkova D., Bloeker H., Scharfe M., Grimm M., Loehner T.-H.,
 RA Dose S., de Haan M., Maarse A.C., Schaefer M., Mueller-Auer S.,
 RA Gabel C., Fuchs M., Fartmann B., Grandrath K., Dauner D., Herzl A.,
 RA Neumann S., Argirou A., Vitale D., Liguori R., Piravandi E.,
 RA Massenet O., Quigley F., Clabaud G., Muendlein A., Felber R.,
 RA Schnabl S., Hiller R., Schmidt W., Lecharny A., Aubourg S.,
 RA Chedof F., Cooke R., Berger C., Monfort A., Cascuberta E.,
 RA Gibbons T., Weber N., Vandenbol M., Barges M., Terol J., Torres A.,
 RA Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse T.,
 RA Heijnen L., Schwarz S., Scholler P., Heber S., Francis P., Bieleke C.,
 RA Frishman D., Haase D., Lemcke K., Mewes H.-W., Stocker S.,
 RA Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Habermann K.,
 RA Parnell L., Dedhia N., Gnoj L., Schutz K., Huang E., Spiegel L.,
 RA Sekhon M., Murray J., Sheet P., Cordes M., Abu-Threideh J.,
 RA Stoneking T., Kalicki J., Graves T., Harmon G., Edwards J.,
 RA Latreille P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,
 RA Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,
 RA Kramer J., Spiehl J., Ryan E., Andrews S., Geisel C., Hillier L.,
 RA Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joshi C.,
 RA Antoniou B., Zidanic M., Strong C., Sun H., Lamar B., Yordan C.,
 RA Ma P., Zhong J., Preston R., Vil D., Shekher M., Matero A., Shah R.,
 RA Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Till S.,
 RA Granat S., Shohdy N., Hasegawa A., Hameed A., Lodhi M., Johnson A.,
 RA Chen E., Marra M., Martienssen R., McCombie W.R.;
 RT "Sequence and analysis of chromosome 4 of the plant Arabidopsis
 thaliana.";
 RL Nature 402:769-777(1999).
 CC -1- FUNCTION: ASPARAGINE-SPECIFIC ENDOPEPTIDASE INVOLVED IN THE
 CC PROCESSING OF VACUOLAR SEED PROTEIN PRECURSORS INTO THE MATURE
 CC FORMS (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Vacuolar.
 CC -1- TISSUE SPECIFICITY: SPECIFIC TO VEGETATIVE ORGANS.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C13.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; D61395; BAA18924.1; ALT_INIT.
 DR EMBL; AL031804; CAA21203.1; ALT_INIT.
 DR EMBL; AL161582; CAB80011.1; ALT_INIT.
 DR MEROPS; C13.002;
 DR InterPro; IPR001096; Legumain.
 DR Pfam; PF01650; Peptidase_C13; 1.
 DR PRINTS; PR00776; HEMOGLOBINASE.
 DR Hydroxylase; Thiol protease; Signal; Multigene family.
 FT SIGNAL 1 25 POTENTIAL.
 FT CHAIN 26 490 VACUOLAR PROCESSING ENZYME, GAMMA-
 FT ISOZYME.
 FT ACT_SITE 173 173 POTENTIAL.
 FT ACT_SITE 215 215 POTENTIAL.
 SQ SEQUENCE 490 AA; 53931 MW; 35A8CA9ABC216DCE CRC64;
 Query Match 50.8%; Score 1246; DB 1; Length 490;
 Best Local Similarity 53.6%; Pred. No. 3.2e-81;
 Matches 233; Conservative 71; Mismatches 121; Indels 10; Gaps 4;
 QY 36 ESSDKSAGTRWALVAGSNYYNRHOADICHAYQILRKGLKDKDENTIVFYDDIAFSS 95
 DB 45 ENDDSSNGTRWALVAGSGYWNRYHOADICHAYQILRKGLKDKDENTIVFYDDIAFSS 104
 QY 96 ENPRGVIINKPDGEDYVYKPKDYKREAVNVQNFYVLLGNESCVTGGNGKVKSGPND 155
 DB 105 ENPRGVIINKPDGEDYVYKPKDYKREAVNVQNFYVLLGNESCVTGGNGKVKSGPND 164
 QY 156 NIFTYADHGAPGIAMPTGDEVMAKDFNEVLEKMKRKKYKMKVYVEACESGSMFEGI 215
 DB 165 HIFYSDHGGPGVLCMPSTPYLVANDLVNKKHALGTYKSLVFLYLEACESGSIPEGL 224
 QY 216 LKKNLNIYAVTAANSKSSWGVCYPPSEIGTCLGDTFSSWLESDSLDHMSKET 275
 DB 225 LPEGLNIYATVTAASNAEESWGTCYPPSEIGTCLGDTFSSWLESDSLDHMSKET 284
 QY 276 LEOQYHVKKR---VGSVDPVTSVHCRCGTEKMLKDYLSYVIGRNPENDNETFTE--SFS 330
 DB 285 LHOQYELVKRTPAPGVSY--GSHVMQYGDVGIKNDNLVYNTNPANDNETFADANSLK 342
 QY 331 SPISNSGLVNRDIPLYLQRIQKAPMGSLKESKAEQKLLDEKKNHRKQIDQSIDILRL 390
 DB 343 PP---SRVTNORDADLVHFWKYEKAPESGARKTEAQVLEAMSHRLHIDNSVILVCKI 399
 QY 391 SVKOTNVNLNLTSTRTTGOPLVDDWDCFKTLVNSFKNHCATVHYGLKYTGALANICNMG 450
 DB 400 LFGISRGPEVLNKRVSAGQPLVDWDCFKTLVNSFKNHCATVHYGLKYTGALANICNMG 459
 QY 451 VDKQTVSAIEQACS 465
 DB 460 IQMEQMEEAQAQACT 474
 RESULT 4
 VPEA_ARATH
 ID VPEA_ARATH STANDARD; PRT; 478 AA.
 AC P49047; 082806;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Vacuolar processing enzyme, alpha-isozyme precursor (EC 3.4.22.-)
 DE (Alpha-VPE)
 GN AT2G25940 OR T19L18.25.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;

RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RX MEDLINE=96017615; PubMed=7579169;
 RA Kinoshita T., Nishimura M., Hara-Nishimura I.;
 RT "Homologues of a vacuolar processing enzyme that are expressed in
 RL different organs in Arabidopsis thaliana.";
 RN Plant Mol. Biol. 29:81-89(1995).
 RP REVISIONS TO 70: 84 AND 91.
 RA Kinoshita T., Nishimura M., Hara-Nishimura I.;
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RX MEDLINE=20083487; PubMed=10617197;
 RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
 RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,
 RA Buell C.R., Ketchum K.A., Lee J.J., Renning C.M., Koo H.L.,
 RA Moffat K.S., Cronin L.A., Shen M., Pai G., Van Aken S., Umayam L.,
 RA Tallon L.J., Gill J.E., Adams M.D., Carrera A.J., Creasy T.H.,
 RA Goodman H.M., Somerville C.R., Copenhaver G.P., Preuss D.,
 RA Nierman W.C., White O., Eisen J.A., Salzberg S.L., Fraser C.M.,
 RA Venter J.C.; and analysis of chromosome 2 of the plant Arabidopsis
 RT thaliana.";
 RL Nature 402:761-768(1999).
 CC -!- FUNCTION: ASPARAGINE-SPECIFIC ENDOPEPTIDASE INVOLVED IN THE
 CC PROCESSING OF VACUOLAR SEED PROTEIN PRECURSORS INTO THE MATURE
 CC FORMS.
 CC -!- SUBCELLULAR LOCATION: Vacuolar.
 CC -!- TISSUE SPECIFICITY: EXPRESSED IN ROSETTE LEAVES, CAULINE LEAVES
 CC AND STEMS. NOT EXPRESSED IN THE SILIQUES.
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C13.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement. Usage by and for commercial
 CC or send an email to license@lsb-sib.ch).
 CC
 CC EMBL: D61393; BAA09614.2; .
 CC EMBL: AC004747; AAC31241.1; .
 CC MEROPS: C13.002; .
 CC InterPro: IPR001096; Legumain.
 CC Pfam: PF01650; Peptidase_C13; 1.
 CC PRINTS: PR00776; HEMOGLOBINASE.
 CC Hydrolase: Thiol protease; Signal: Multigene family.
 CC SIGNAL 1 20 POTENTIAL.
 CC CHAIN 21 478 VACUOLAR PROCESSING ENZYME, ALPHA-
 CC ACT_SITE 162 162 ISOZYME
 CC ACT_SITE 204 204 POTENTIAL.
 CC SEQUENCE 478 AA; 52670 MW; D73DIF353E2FE898 CRC64;
 Query Match 50.1%; Score 1230.5; DB 1; Length 478;
 Best Local Similarity 50.7%; Pred. No. 3.9e-80;
 Matches 233; Conservative 79; Mismatches 137; Indels 11; Gaps 5;
 QY 10 ILVFLHALLIFSASRRKTLQNDNDVSSDKSAGKTRWAVLVAGSNYYNRHQADICHA 69
 DB 8 LALFLFLVAASGVDTKLPLSLASKFRPTENDDDSTKAVLVAGSSGYNYNRHQADVCHA 67
 QY 70 YQLRKGKGLDENIIVFYMDIAFSSSNRPVGIINKPDGVDYKGVPKDYTKKAVNVQN 129
 DB 68 YQLKGGVKEENIVFVMDYDIKAKNEENRPVGIINSPNGEDVYNGVPKDYTGDEVNVDN 127
 QY 130 FYNVLGNESGVGGNGKVVVKGSPNDNIFYADHGCAPGLIAMPTGDEVNAKDFNEVLK 189
 DB 128 LLAVLGNKTKALKGSGKVVVDSGPNDFIYIYSDHGGPGVGLMPTSPNLVANDLNDLKK 187

QY 190 MHKRRKYNKVIYVEACSGSMFEGILKKLNINIVYTAANSKSSWGVCYSPESYPPPPSE 249
 DB 188 KYASTGYKSLVFLYLEACSGSIFGELLPEGLNIYATTAASNAESSWGTCYCGEDPPSPSE 247
 QY 250 IGTCLGDTFISWLESDLDHMSKETLEQQYHVYKRR-VGSDVPETSHVCRFGCTERMLKD 308
 DB 248 YETCLGDLYSVAMIEDSEKHNLQOTETLHEQYELVKRTAGSGKSYGSHVMEFGDIGLSKE 307
 QY 309 YLSSVYIGRNPENDNFTFT--ESFSPISNSGLVNPRIPLLYLQRIKQKAPMGSLSKSEA 366
 DB 308 KLVLFMGTPADENFTFVNENSIRPP---SRVTTNQRDADLVHFWHKYQKAPGEGSARKVEA 364
 QY 367 OKLLDEKNHRKQIDQSI--TDILRLSVKQTNVNLNLTSTRTTGOPLVDWDDCFKTLVNS 424
 DB 365 OKQVLEAMSHRLHVDNSILLIGLLFGLGHAFLN---KVRPSGEPLVDWDDCLKSLVRA 421
 QY 425 FKNHCGATVHYHLYGTGALANIKMGVDVQVTSALIEQAC 464
 DB 422 FERHCGSLSOYGIKHMRSIANCMNAGIQMROMEEAQAQAC 461
 RESULT 5
 VPEB_ARATH STANDARD; PRT; 486 AA.
 AC Q39044; Q9SI79; Q93VS7;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DE Vacuolar processing enzyme, beta-isozyme precursor (EC 3.4.22.-)
 DE (Beta-VPE).
 GN ATIG62710 OR F23N19.7.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OC NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RX MEDLINE=96017615; PubMed=7579169;
 RA Kinoshita T., Nishimura M., Hara-Nishimura I.;
 RT "Homologues of a vacuolar processing enzyme that are expressed in
 RL different organs in Arabidopsis thaliana.";
 RL Plant Mol. Biol. 29:81-89(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RX MEDLINE=21016719; PubMed=11130712;
 RA Theologis A., Ecker J.R., Altafi H., Araujo R., Bowman C.L., Brooks S.Y.,
 RA White O., Alonso J., Chao Q., Chen H., Cheuk R.E., Cressy T.H., Dewar K.,
 RA Buehler E., Chan A., Chao Q., Conway A.B., Conway A.R., Fong B., Fujii C.Y.,
 RA Chung M.K., Conn L., Conway A.B., Feng J.-D., Fong B., Huizar L.,
 RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
 RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
 RA Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
 RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
 RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziani A.,
 RA Militscher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
 RA Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
 RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
 RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
 RA Uterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
 RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
 RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis
 RL thaliana.";
 RL Nature 408:816-820(2000).
 RP [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RA Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;
 RT "RIKEN Arabidopsis full length cDNA clones (RAFLs) sequenced by the
 RT SSP consortium (Saik/Stanford/PCEC).";

QY 275 TLEOQHVVKRRYV--SDVPETSHVCRFTEKMLDYLSYIGRNPNENDNFTTETESFSP 332
 Db 291 TVAQOYSSVKARTSNYNTAAAGSHVWQYQNSIKADKLYLFOGFOFASVNF-----P 342
 QY 333 ISNSGL-----VNPDPILLYLORKIOPKAPGCSLESKEAKKLLDKNRHKKQIDOSTD 386
 Db 343 PNAHLNAPMEVYNORDAEHFWWQLYKSENGSEKKELIOIKDAIKHRSHLDSSWOL 402
 QY 387 ILRLSVKQTNVLLNLTSTRTTQPLVDDWDCFKTLVNSFKNHCIGATVHYGLKYTGALANI 446
 Db 403 IGDLFGPKKASALKSVREPGSPLVDDWGLKSVWVFPETCCGSLTOYGMKHMRTFANI 462
 QY 447 CNMGVDKQTVSAIEOACS 465
 Db 463 CNAGVSHTSMEEAACNAACS 481

RESULT 7
 LEGU_CANEN STANDARD; PRT: 475 AA.
 AC P49046;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DE Legumain precursor (EC 3.4.22.34) (Asparaginyl endopeptidase).
 OS Canavalia ensiformis (Jack bean) (Horse bean).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Canavalia.
 NCBI_TaxID=3823;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Seed;
 RX MEDLINE=95155263; PubMed=7852272;
 RA Takeda O., Miura Y., Mitta M., Matsushita H., Kato I., Abe Y.,
 RA Yokosawa H., Ishii S.;
 RT Isolation and analysis of cDNA encoding a precursor of Canavalia
 RT ensiformis asparaginyl endopeptidase (legumain).;
 RL J. Biochem. 116:341-346(1994).
 RN [2]
 RP SEQUENCE OF 36-60. AND CHARACTERIZATION.
 RX MEDLINE=93155205; PubMed=8429028;
 RA Abe Y., Shirane K., Yokosawa H., Matsushita H., Mitta M.,
 RA Kato I., Ishii S.;
 RT Asparaginyl endopeptidase of jack bean seeds. Purification,
 RT characterization, and high utility in protein sequence analysis.;;
 RL J. Biol. Chem. 268:3525-3529(1993).
 RN [3]
 RP REVIEW.
 RX MEDLINE=95147717; PubMed=7845236;
 RA Ishii S.;
 RT Legumain: asparaginyl endopeptidase.;;
 RL Meth. Enzymol. 244:604-615(1994).
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of proteins and small-molecule
 CC substrates at -Asn|-Xaa-bonds.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C13.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC o. Send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: D31787; BAA06596.1; -
 CC InterPro: IPR001096; Legumain.
 CC Pfam: PF01650; Peptidase_C13; 1.
 CC PRINTS: PR00776; HEMOGLOBINASE.
 CC HydroLase: Thiol protease; Signal; Zymogen.
 CC SIGNAL 1 15 POTENTIAL.

FT PROPEP 16 35
 FT CHAIN 36 475 LEGUMAIN
 FT CARBOHYD 300 300 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT ACT_SITE 158 158 POTENTIAL.
 FT ACT_SITE 200 200 POTENTIAL.
 SQ SEQUENCE 475 AA; 52763 MW; 6D1C1D6872C5504C CRC64;
 Query Match 44.0%; Score 1080; DB 1; Length 475;
 Best Local Similarity 48.3%; Pred. No. 1.9e-69;
 Matches 207; Conservative 73; Mismatches 129; Indels 20; Gaps 4;
 QY 44 GTRMAVLVAGSNEYNNRHOADICHAYQILRKGGLKDENIIVFMYDDIIAFSSNPRPGVI 103
 Db 38 GTRMAVLVAGSNGYGRHOADYCHAYQLLIKGVKEENIVFMYDDIIAYNANPRPGVI 97
 QY 104 INKPDGEDVYKGVPKDYTKRANVQNFYVNLGLNESGVTGGNGKVKVSGPNDFIYAD 163
 Db 98 INHPQGPVYAGVPKDYTGEDVTPENLYAVILGDKSVKVGSGKVINSPEDRIIFYSD 157
 QY 164 HGAPGLIAMPPTGDEVMAKDFNEVLEKMKHKKYKNKVIYVEACESCMEFEGILKKNLIY 223
 Db 158 HGGPGVLGMPNAPFVAMDFIDVLKHKHAGSGYKENVIVIEACESGIFEGIMPKDLNIY 217
 QY 224 AVTAANSKESWGVCYCPSPPPPPSIGTCLGDTFISWLESDSLHDMSKETLEQOYHV 283
 Db 218 VTTASNAOENSFCTYCPGMNPPPEEVVTCGLDLYSVSNMEDSETHNLKRETVOQOYQSV 277
 QY 284 KRRVGSVPET--SHVCRFGTEKMLDYLSYIGRNPNENDNFTTETESFSPISNSGL--- 338
 Db 278 RKRTSNNSYRFGSHVNOYGDNTITAEKLYLYHGPDATVNF-----PPHNGNLEAK 329
 QY 339 ---VNPDPILLYLORKIOPKAPGCSLESKEAKKLLDKNRHKKQIDOSTDITLRLSVKT 395
 Db 330 MEVVNQDAELLPWQYQNSNQHPKTHILEQITETVKNRHLDCSVELGLVLYGPG 389
 QY 396 NVLLNLTSTRTTQPLVDDWDCFKTLVNSFKNHCIGATVHYGLKYTGALANICNMGVDDVKQ 455
 Db 390 KSSVLSHVRAPGLPLVDDWTCUKSVRVFETHCGSLTOYGMKHMRAFQNCVSGV---- 445
 QY 456 TVSAIEOAC 464
 Db 446 SKASMEEAC 454

RESULT 8
 VPE_SOYBN STANDARD; PRT: 495 AA.
 ID VPE_SOYBN
 AC P49045;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DE Vacuolar processing enzyme precursor (EC 3.4.22.-) (VPE).
 OS Glycine max (Soybean).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
 NCBI_TaxID=3847;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Seed cotyledon;
 RX MEDLINE=94356350; PubMed=8075902;
 RA Shimada T., Hiraiwa N., Nishimura M., Hara-Nishimura I.;
 RT Vacuolar processing enzyme of soybean that converts proproteins to
 RT the corresponding mature forms.;;
 RL Plant Cell Physiol. 35:713-718(1994).
 CC -1- FUNCTION: ASPARAGINYL-SPECIFIC ENDOPEPTIDASE INVOLVED IN THE
 CC PROCESSING OF VACUOLAR SEED PROTEIN PRECURSORS INTO THE MATURE
 CC FORMS.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C13.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC o. Send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: D31787; BAA06596.1; -
 CC InterPro: IPR001096; Legumain.
 CC Pfam: PF01650; Peptidase_C13; 1.
 CC PRINTS: PR00776; HEMOGLOBINASE.
 CC HydroLase: Thiol protease; Signal; Zymogen.
 CC SIGNAL 1 15 POTENTIAL.

QY 158 FIYADHAGPLIAMPTGDEVNAKDEFNEVLEKMKHKKYKMYIYVACESGSMFEGILK 217
 DB 144 FIYFDHAGTGLVFP--NDLHVKDLNKTIRYMEHKYQRMVYIEACESGSMNH--LP 201
 QY 218 KNLNIYAVTAANSKSSMGVYCPSPPPSEIGTCLGDTFISWLESDLDHMSKETLE 277
 DB 202 DDINIVYATTAANKSSYACYDE-----ERGTGLDWTYSVMNEDSDVEDLTKETLH 254
 QY 278 QYHVYKRVGSDVPSHVCFRGTGTEKMLKDYLSYIGRNPNENDNFTTFSFSPIS--- 334
 DB 255 KOYHLVKSHT-----NTSHVMQYGNKSISTMKVMQFQ-----MKHRASSPISLPP 300
 QY 335 --NSGLVNRPRDIPLLYLORKIKAPMGSLSEKQ-----KLLDEKNHRKQIDQSTTD 386
 DB 301 VTHLDLTPSPDVPPLTLTKRKL-----LRTNDVKESQNLIGIQIQELDA---RHVIEKSVHK 353
 QY 387 IRLSVKOTNVLNLLTSTRTTQGLVDDWDCFKTLVNSFKNHC-----CATVHYGLKYTGA 442
 DB 354 IVSLLAGFGETAERHLSERT-----MLTARDCYQEAUTHERTHCFNWHSVTYEHALRYLYV 409
 QY 443 LANICNMGVDVKTVSATBOAC 464
 DB 410 LANLCEAPYPIDRIEMADKVC 431

RESULT 10
 LGMN_RAT
 ID LGMN_RAT STANDARD: PRT: 435 AA.
 AC Q9R008; Q9JL3;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Legumain precursor (EC 3.4.22.34) (Asparaginyl endopeptidase)
 DE (Protease, cysteine 1).
 GN LGMN OR PRSC1.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RC TISSUE=Liver;
 RP SEQUENCE FROM N.A.
 RA Ishidoh K.;
 RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Neumann J., Koehler B., Reske K.;
 RT "Cloning and expression of rat legumain."
 RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: Has a strict specificity for hydrolysis of asparaginyl
 CC bonds. Can also cleave aspartyl bonds slowly, especially under
 CC acidic conditions. May be involved in the processing of proteins
 CC for MHC class II antigen presentation in the lysosomal/endosomal
 CC system (By similarity).
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of proteins and small-molecule
 CC substrates at -Asn-[Xaa]-bonds.
 CC -!- SUBCELLULAR LOCATION: Lysosomal (By similarity).
 CC -!- PTM: Glycosylated (Probable).
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C13.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: AB032766; BAA84750.1;
 CC EMBL: AF154349; AAF73260.1;
 CC MEROPS: C13.004;
 CC InterPro: IPR001096; Legumain.
 CC Pfam: PF01650; Peptidase_C13; 1.

DR PRINTS: PR00776; HEMOGLOBINASE.
 KW Hydroxylase; Thiol protease; Lysosome; Glycoprotein; Zymogen; Signal.
 FT SIGNAL 1 17
 FT CHAIN 18 325
 FT PROPEP 326 435
 FT SITE 325 326
 FT CARBOHYD 93 93
 FT CARBOHYD 169 169
 FT CARBOHYD 215 215
 FT CARBOHYD 265 265
 FT CARBOHYD 274 274
 FT ACT_SITE 150 150
 FT ACT_SITE 191 191
 FT CONFLICT 2
 SQ SEQUENCE 435 AA; 49466 MW; E835F21C131F7A98 CRC64;

Query Match 33.9%; Score 832; DB 1; Length 435;
 Best Local Similarity 40.2%; Pred. No. 7.3e-52;
 Matches 180; Conservative 75; Mismatches 131; Indels 62; Gaps 13;

QY 39 DKSAGTRWAVLVAGSNYYNYRHQADICHAYQILRKGLADENIIVMYDDIAFSSEN 98
 DB 24 DPEDGKHWVIVAGSNWYNYRHQADACHATQIIHRNGIPDEQLIVMYDDIANNEEN 83
 QY 99 RGVIIINPKDGEDYKGVKPYKDYTKAVNVQNFYVLLNKGSGVTG--GNGKVVKSGPDNI 157
 DB 84 TPGVIVNRPNGTGVYKGVKPYKDYTGEDVTPENFLAVLRGDEEAVKVGSGKVLKSGPRDHV 143
 QY 158 FIYADHAGPLIAMPTGDEVNAKDEFNEVLEKMKHKKYKMYIYVACESGSMFEGILK 217
 DB 144 FIYFDHAGTGLVFP--NDLHVKDLNKTIRYMEHKYQRMVYIEACESGSMNH--LP 201
 QY 218 KNLNIYAVTAANSKSSMGVYCPSPPPSEIGTCLGDTFISWLESDLDHMSKETLE 277
 DB 202 DDIDVYATTAANKSSYACYDE-----ERSTYLGDMYSVMNEDSDVEDLTKETLH 254
 QY 278 QYHVYKRVGSDVPSHVCFRGTGTEKMLKDYLSYIGRNPNENDNFTTFSFSPIS--- 334
 DB 255 KOYHLVKSHT-----NTSHVMQYGNKSISTMKVMQFQ-----MKHRASSPISLPP 300
 QY 335 --NSGLVNRPRDIPLLYLORKIKAPMGSLSEKQ-----KLLDEKNHRKQIDQSTTD 386
 DB 301 VTHLDLTPSPDVPPLTLTKRKL-----LRTNNKESQVLVQGIQHLDDARH----- 345
 QY 387 IRLSVKOTNVLNLLTSTRTTQGLVDDWDCFKTLVNSFKNHC-----CATVHYG 436
 DB 346 ILEKSVQK--IVSLLAGFGETAQKHLSEKRLTAHDCQEAUTHERTHCFNWHSVTYEHA 403
 QY 437 LKATGALANICNMGVDVKTVSATBOAC 464
 DB 404 LRLYVLANLCEAPYPIDRIEMADKVC 431

RESULT 11
 LGMN_HUMAN
 ID LGMN_HUMAN STANDARD: PRT: 433 AA.
 AC Q99538; O00123;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Legumain precursor (EC 3.4.22.34) (Asparaginyl endopeptidase)
 DE (Protease, cysteine 1).
 GN LGMN OR PRSC1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Placenta;
 RX MEDLINE=97218252; PubMed=9065484;
 RA Chen J.-M., Dando P.H., Rawlings N.D., Brown M.A., Young N.E.,
 RA Stevens R.A.E., Hewitt E., Watts C., Barrett A.J.;


```

Db 277 LYLSEFGQSRKK-----ASTERDEPPMKPKDSIPSRDPLHTLHRRIMMA--NNMNDKTL 330
QY 368 KLLDKNHRKQI---DQSTIDILRLSVKQTNVNLTLSTRTTQOPVDDWDCFKTLVNS 424
Db 331 MKILGLKLRDLKDTMEVIDQFMFNKPN-----SNATIDETM-----DCIEVYKE 380
QY 425 FKNHCGATVHYGLKYTGALA---NICNMGVDDVKOTVSAIEQAC 464
Db 381 FOSKC-FKIQOAPETITGLVTLNYCQKGYSAENINGVIRKVC 422

RESULT 14
ID GPI8_YEAST STANDARD; PRT; 411 AA.
AC P49018;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE GPI-anchor transamidase (EC 3.-.-.-) (GPI transamidase).
GN GPI8 OR YDR331W OR D9798.2.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288c / AB972;
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,
RA Favetto A., Fulton L., Gattung S., Greco F., Kirsten J.,
RA Kucaba T., Hallsworth K., Hawkins J., Hillier L., Jier M.,
RA Johnson D., Johnston L., Langston Y., Latreille P., Le T.,
RA Mardis E., Menezes S., Miller N., Nhan M., Pauley A., Peluso D.,
RA Rifken L., Riles L., Taich A., Trevasakis E., Vignati D.,
RA Wilcox L., Wohldman P., Vaudin M., Wilson R., Waterston R.;
RA Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP CHARACTERIZATION.
RX MEDLINE=97133287; PubMed=8978684;
RA Benghezal M., Benachour A., Rusconi S., Aebi M., Conzelmann A.;
RT "Yeast Gpi8p is essential for GPI anchor attachment onto proteins.";
RL EMBO J. 15:6575-6583(1996).
CC -!- FUNCTION: Mediates GPI anchoring in the endoplasmic reticulum, by
CC replacing a protein's C-terminal GPI attachment signal peptide
CC with a pre-assembled GPI. During this transamidation reaction, the
CC GPI transamidase forms a carbonyl intermediate with the substrate
CC protein.
CC -!- PATHWAY: GPI-anchor biosynthesis.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C13.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U32517; AAB64766.1;
CC MEGR08P; C13.005;
CC SCD; S0002739; Gpi8.
CC InterPro: IPR001096; Legumain.
CC Pfam: PF01650; Peptidase_C13; 1.
CC PRINTS; PR00776; HEMOGLOBINASE.
CC Hydrolase; Thiol protease; GPI-anchor biosynthesis.
FT ACT_SITE 157 157 POTENTIAL.
FT ACT_SITE 199 199 POTENTIAL.
SQ SEQUENCE 411 AA; ED9F4A1C3B214E28 CRC64;

Query Match 13.1%; Score 321; DB 1; Length 411;
Best Local Similarity 25.1%; Pred. No. 1.4e-15;
Matches 115; Conservative 77; Mismatches 169; Indels 98; Gaps 18;

```

```

QY 7 HFQILVELHALLIFSAESRRKTQLLNDNDVSSDKSAKTRMAVLVAGSNEYNYNRHOADI 66
Db 6 HLPULLLYIFLLPLSG-----ANNTDAAHEVIATNTNNMAVLVSTSRWFENYRHMANV 58
QY 67 CHAYQILRKGLKDKDENILVPMYDIAFSSSENPRGVIIKPD-GEVY-KGVPKDYTKEA 124
Db 59 LSMYRTVKRLGIPDSQIILMLSDDVACNSRNLFPGSVFNKNDHAIDLGYDSVEVDYRGYE 118
QY 125 VVONFYNVLLGNESGVTTGGNGKVKVSGPNDIFYYVADHGCAPGLIAMPTGDEVNAKDFN 184
Db 119 VTVENFIR-LLTDRWTEDEHPKSKRLTLTDENSNFIYVTGHHGDDFLKFDQAEETASEDIA 177
QY 185 EVLEKMHKRYKKNMVIYVEACSGSMFEGILKKNLNIYAVTAANSKSSGNGVYCPESYP 244
Db 178 DAFQOMEKKRYNEIFPMIDTQANTWYSKFYSP--NILAVGSEMDESSYSHH-----229
QY 245 PPSEIGTCGLDFTFS-----ISWLEDS-----LHDMSKETLEQOYHVVKRVSQVDPETS 295
Db 230 -SDVEIGVAVIDRFTYYCCLDFLEQIDKNSTLTLODLFDSFTFEKIH-----S 275
QY 296 HV-----CRFGTEKMLKDYLSYIGRNENDNFTTESFSSPISGLVNPRIPLLY 348
Db 276 HVGVRTDLFRNPSEVLITDEFANQVNPIDD-----SKPLSVSHYHYKD-----321
QY 349 LQRKIQAPMGSLSKQAQKLLD-----EKNHRKQIDQISITDILRLSVKOTNVNL 401
Db 322 ---HIDTA-----QYELNNVLDLALETYKKNOSKIEKKIKDI-----KTSVLVDV- 366
QY 402 TSPRTQOPLVDDWDCFKTLVNSKHCATVHYGLKYT 440
Db 367 -----DIDSNECFPT---SFKQ---SATIILALIVT 391

RESULT 15
GPI8_HUMAN STANDARD; PRT; 395 AA.
ID Q92643; O14822;
AC Q92643; O14822;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE GPI-anchor transamidase (EC 3.-.-.-) (GPI transamidase)
DE (phosphatidylinositol-glycan biosynthesis, class K protein) (PIG-K)
DE (HGP18).
GN PIGK OR GPI8.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97133287; PubMed=8978684;
RA Benghezal M., Benachour A., Rusconi S., Aebi M., Conzelmann A.;
RT "Yeast Gpi8p is essential for GPI anchor attachment onto proteins.";
RL EMBO J. 15:6575-6583(1996).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=98024174; PubMed=9356492;
RA Yu J., Nagarajan S., Knez J.J., Udenfriend S., Chen R., Medof M.E.;
RT "The affected gene underlying the class K glycosylphosphatidylinositol
RT (GPI) surface protein defect codes for the GPI transamidase.";
RL Proc. Natl. Acad. Sci. U.S.A. 94:12580-12585(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA Strausberg R.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Mediates GPI anchoring in the endoplasmic reticulum, by
CC replacing a protein's C-terminal GPI attachment signal peptide
CC with a pre-assembled GPI. During this transamidation reaction, the
CC GPI transamidase forms a carbonyl intermediate with the substrate
CC protein.
CC -!- PATHWAY: GPI-anchor biosynthesis.
CC -!- SUBUNIT: Associates with PIGS and PIGT.

```

```

CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C13.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: Y07596; CAA68871.1;
CC EMBL: AF022913; AAB81597.1;
CC EMBL: BC020737; AAB81597.1;
CC Genew: HGNC:8965; PIGK.
CC MIM: 605087;
CC MEROPS: C13.005;
CC InterPro: IPR001096; Legumain.
CC Pfam: PF01650; Peptidase_C13; 1.
CC PRINTS: PR00776; HEMOGLOBINASE.
CC Hydrolase: Thiol protease; GPI-anchor biosynthesis.
FT ACT_SITE 164 164
FT ACT_SITE 206 206
FT CONFLICT 1 4
FT CONFLICT 4
SQ SEQUENCE 395 AA: 45251 MW: 4706 DDDADDD13EFB2 CRC64;

Query Match 12.1%; Score 297.5; DB 1; Length 395;
Best Local Similarity 23.4%; Pred. No. 6.4e-14;
Matches 95; Conservative 82; Mismatches 170; Indels 59; Gaps 14;

QY 14 LHALFSAESRTQLLNDNDVSSDKSAGTRWAVLVAGSNEYNYRHRQADICHAYOIL 73
DB 14 LATVLLSFGSVAASHIED-QAEQFFRSCHTNNWAVLVCTSRWFNYRHRVANTLSVYRSV 72

QY 74 RKGGLDENIIVFYDDIAFSSNPRPGVIINKPDGE-DVY-KGVPKDYTKYKAVNVQNFY 131
DB 73 KRLGIPDSHIVLMLADDMACPNRPKPATVFSKKNMELNVYGDDEVDYRSYEVTVENFL 132

QY 132 NVLLGNESGVGTGGNGKVKVSGPNDFIYYADHGAPGLIAMPDTGDEVMAKDFNEVLEKMH 191
DB 133 RVLTGRIPPST-PRSKRLLSDDRSNLIYMTGGNGFLKFDQSEITNIELADAFQOMW 191

QY 192 KKKYKMKVIYVEACSGSMFEGILKKNLNIYAVTAANSKSSWGVCYCPESYPPPPSEIG 251
DB 192 QKRYNELLFIIITCOGASWYERFSP--NIMALASSQVGEDSL-----SHOPDPA-IG 242

QY 252 TGLGDTFSISWLED-SDLHDMKRETLQOYHVVKRVGSDVPETSHVCRFGTEKMLKDYL 310
DB 243 VHLMDRYTFVVFLEFEINPASOTNNNDLFQVCPKSLCVSTP-----GHR 287

QY 311 SSYIGRNPNENDFTFTESPSSPISNSGLYNPRDIPLLYLQRIKAPWGSLESKEAKKL 370
DB 288 TDLFQDPK--NVLIITDFGS-----VRKVEITTIKLOQDSE 324

QY 371 LDEKNHRK-QIDQSIITDILRLSVKQTNVNLITSTTTGQPLVDDW 415
DB 325 IMESSYEDQMDERKLEPLKYA-EQLPVAQIIHQ-----KPKLQDW 364

```

Search completed: May 27, 2003, 15:26:51
Job time : 17 secs